Intentional Agency and Temporal Structures of Imputation

Sebastián Figueroa Rubio (Universidad Autónoma de Madrid)

Abstract. As a rule, the study of responsibility for actions adopts an episodic view of agency, according to which, each action is defined as a single event that breaks a norm. Furthermore, from this view, an action is intentional under a description that shows that has been caused by certain mental states. Nevertheless, our accountability practices sometimes are based on a temporally extended conception of agency to which we are responsible for a set of actions that are temporally extended and encompass a variety of events (e.g., *action libera in causa*). Moreover, we assume that some wrongs are intentional even when no mental state caused them (e.g., negligence). In this text, I examine some structures that assumes this temporally extended conception of agency and their connections to intentional action. To this end, I analyse some ideas developed by Michael Bratman on agency and by Cristina Bicchieri on how we interpret the actions of others, and I explore how these ideas can help shed light on how some legal categories operate.

I

It is common to think that being a free agent is a necessary condition for being responsible. What this actually means, and how it is reflected in our judgements of attribution of responsibility, is a topic that has been widely discussed. In this presentation I focus on how thinking about the temporal structure of our practical experience makes it possible to say some things about the relation between agency and responsibility, and how this is reflected in some legal institutions.

First, I would like to note that the concept of free agency has a variety of aspects, not all of which are related to responsibility in the same way. Carlos Moya (2017 ch. 1), who deals with this subject from the perspective of control, shows many of these aspects. There are three of them that are important to consider for the purposes of this presentation. The first is volitional control. According to this, agency implies the possibility of acting voluntarily and intentionally. This means that the agent's behaviour is goal-directed and non-accidental. This implies the ability to

represent goals and to develop the kind of control that enables agents to carry them out more or less successfully. The second is rational control. According to this notion, the exercise of agency can be done for reasons, and agents are both sensitive to reasons and able to give reasons to explain, justify, and shape their behaviour. Finally, there is origin control, meaning that the reasons and intentions of an agent can express her deep commitments and perspectives. Thus, sometimes agent's behaviour expresses her "fundamental evaluative orientation," which shows her "practical identity". (Watson, 1996: 271).

From these three aspects, we can see that agency involves not only voluntary action, but also a set of cognitive skills, the ability to interact with others, and the development of a self-image. The exercise of agency expresses much more than simply bringing about changes in the physical world. In this context, the concept of intention is closely related to concepts such as action, mind, and responsibility, and usually acts as a link between them. Intentional actions are a part of our behaviour that is under our control, and at the same time they relate to our commitments, playing a dual role as they are linked to our sensory-motor abilities, on the one hand and to our social-normative life on the other. Against this background, to characterize other's actions as intentional “provide a basis for our everyday attempts to predict what others will do, explain what they done, and coordinate our projects with others” (Bratman, 1999:1: v. Anscombe, 1963: par 1). Thinking about this has led to the development of some critiques of the prevailing theory of action.

The prevailing theory focuses on understanding action as causing changes in the world voluntarily. In his seminal article *Agency*, Donald Davidson wrote that "we never do more than move our bodies" (Davidson, 1980: 59) expressing that these movements, which he called primitive actions (1980: 49), are what really count as the exercise of our agency. The other things that are relevant to the language of action are events that surround and are causally connected to these body movements.

Along these lines, philosophers of action have generally directed their interest to the study of more or less simple and concrete actions and the mental states that accompany them. For example, some mental states, such as desires and beliefs, are defined as causes of actions (Davidson, 1980. Ch.1).

In this context it is common to see intentions as the agent's representation of certain immediate effects that she wants to produce in the world (Davidson, 1980: ch. 5). This leads to a philosophy of action based on the concept of "proximal" intentions, that is, intentions about what to do now (Mele, 2019). These mental states link their actions to the consequences of the movements of our bodies and open the possibility of considering some events as part of the correct description of the agent's action. In this model, each action is defined and separated from each other by these causal relations, creating an episodic view of our agency.

II

This model has been criticised by Michael Bratman, whose ideas I would like to explore now. First, Bratman proposes to make the understanding of the intentional element of our agency more complex by suggesting that intentions refer not only to what we do now, but also to future actions as well as to more or less complex activities. In this way, he suggests including the notion of plan to our vocabulary of intention. For him, this represents a change in our understanding of agents, for it is no longer just about who voluntarily and intentionally cause changes in the world, but also about who plan what they do.

One consequence of this is the need to revise the model that focuses on the desire-belief pair to explain actions. Thus, in his article *Two Faces of Intention* he writes:

> I think this approach is mistaken. We are planning creatures. We frequently settle in advance for the future. On occasion, this even involves settling on one of several conflicting options each of which is, in light of our desires and beliefs, equally attractive. These plans help guide our later conduct and coordinate our activities over time, in ways in which our ordinary desires and beliefs do not. Intentions are typically elements in such coordinating plans. Once we recognize this central role of intentions play in our lives the natural view to take, I think, is that intentions are distinctive states of mind, not to be reduced to clusters of desires and beliefs (1984: 376).

Hence, it is important is to understand intentions as a specific mental state, different from desires and beliefs. Thus, unlike desires, intentions are subject to the rational pressure of consistency.

Additionally, the belief in something future does not imply, as in the case of intentions, the determination of certain means to achieve it. Intentions are also different from predictions, because when we intend to make something happen, we not only refer to possible future events, but we also adopt a commitment towards them. Intentions show aspects of our agency in its own terms.

Regarding plans, we can say that while they are not completely independent of desires and beliefs, they are in some sense different from them and may conflict with them. According to Bratman, one can identify a kind of normativity inherent in plans, which assumes that they function as reasons that can be imposed on concrete desires and beliefs, as well as resolve conflicts between them. Moreover, plans also incorporate "distal" intentions within the picture, that is, intentions that take into account future actions and more or less complex processes that extend over time and of which we are a part. As the author writes: "plans are not merely executed. They are formed, reconsidered, combined, constrained by other plans, filled in, modified and so on" (1984: 89)

Bratman's proposal has two important implications. The first is that there may be conflicts between proximal intentions as well as between them and our plans, and plans help to resolve such conflicts and to interpret an action as part of a whole series of activities. The second implication is that actions are not simply episodes in an agent's life but are embedded in a complex flow of life. The latter, in turn, makes it possible to see the various connections between different actions over time, as well as between an agent's reasons. Thus, agents' activity is no longer defined as a series of episodes that follow one another, but as a complex temporal unit for which the study of reasons and rationality plays an important role.

An example of this idea is provided by Luca Ferrero (2017) with what can be called the Course of Rational Active Intelligent Guidance" (CRAIG), understood as individual continuous processes, such as playing the piano. According to Ferraro, "once the agent S acquires a goal G, she thereby begins *to see to it* that G". The activity in question is not just the effect of an intention, but "It is the outcome of the agent's ability to shape her conduct out of her appreciation of the demands that the pursuit of G imposes at any given time in light of her contemporaneous view of her present

and future circumstances, opportunities, skills, abilities, and goals." (2017)[1]. As Bratman points out:

> In many cases of human agency the agent's present activity involves her grasp of how it is embedded in what she has earlier been doing and what she is on her way to doing. Her grasp of the larger temporal arc of her activity is a central element in her guidance of that activity both at that time and over time. And this grasp is central to our understanding of her activity". (2010: 8)

The notion of plan, then, enables us to give meaning to many of our actions and to generate different descriptions of our doings. For example, drinking water can simply mean getting hydrated, but in certain contexts, such as a football game, it can be understood as a way to buy time in order to influence the course of the game and thus defeat the opponent. As Bratman writes: "We typically see present actions as elements in planned activities that extend over time. Frequently, it is only when is seen in this light that our present activities make the right kind of sense to us" (1999: 165)

III

In view of what has been said so far, intentional action turns out to be an important part of what is involved in exercising responsible agency. I would now like to dwell briefly on what intentional action means. To begin with, and following Wilfrid Sellars (1980), we can distinguish three elements that are present in the formation of an intention. First, there are some beliefs of the agent about the factual world that enter her practical reasoning as aspects of the context that condition what can be done. This can be distinguished from what depends on what the agent is able to perform within her capacities, what is up- to her in order to achieve what is intended. In the football example, the former is reflected in the time left until the end of the game, while the latter is what depends on the time the player gains, such as the time it takes her to drink the water. In addition, there is a third group of elements that correspond to contextual facts that, although not part of what is intended, can change the meaning of what is happening. In the expression or attribution of an

[1] In this sense, it integrates different levels of intentionality (v. Gallagher, 2020: ch. 3)

intention, whether distal or proximal, these three elements are present whether or not the agent is aware of them, and whether or not they are made explicit by her reports or counterfactually by information to which others have access.

To this it should be added that the possibility of consciously adopting our intentions, of reporting them, and of others being able to refer to them, allows us to recognise certain temporal elements inherent in them. This can be articulated in various ways. On the one hand, by explaining what we are doing or trying to do, so that when a teammate asks the player what Is she doing, she can respond, "I am buying us time." On the other hand, reports can also be made before performing the relevant actions by indicating what you intend to do in the future. So, if someone asks me what I intend to do tomorrow, I can answer, "I am going to play football." Finally, when the relevant events have occurred, what has been done can be reported by creating descriptions that refer to the intention with which it was done. In this way, our understanding of the exercise of intentional agency is embedded in the continuum of our practical lives and relates to what is happening, what will happen, and what has happened. In this way it is possible to develop narratives in which we appear as actors in what happens and through which our behaviour and that of others is explained.

These features of intention can be enriched within the planning agency framework. Thus, according to Bratman:

> Intentions are plan states. Though they are subject to revision, these plan states nevertheless have a characteristic stability over time. And plan states normally adjust in the direction of intention-belief consistency and means-end coherence of plans at a time. They are responsive to pressures for consistency of the many different things one intends with other and with what one believes; and they are responsive to pressures fill in hierarchically structured partial plans as needed with specifications of means and the like. (2010: 9)

Therefore, in addition to the three elements Sellars identifies, Bratman notes that intentions help agents to develop diachronic stability and synchronic consistency. Acting intentionally has a snowball effect because when we act on a plan, we move closer to achieve that plan and further away from achieving another. When we embrace a plan and begin to act in accordance with it, many of our proximal intentions align with it. This is related to the fact that reconsidering a plan

requires time and the use of various mental resources, which can be very costly. Moreover, this snowball effect leads us to acquire habits and other non-deliberative strategies that become part of our behaviour. All this creates a certain rational pressure towards consistency between our goals and the means to achieve them on the one hand, and between different plans and intentions on the other.

The above, in turn, shows how the various types of control mentioned at the beginning interact to form practical agency. On the one hand, in many cases actions can be manifestations of one kind of control but not the other. For example, certain intentional actions are not necessarily part of a plan, nor do they manifest an agent's deep commitments, even if they are voluntary. In turn, when interpreting actions, it may be the case that such commitments manifest themselves beyond the conscious awareness of the agent when she performs an action. In this sense, actions may be described differently by the agent and by others, depending on which perspective one takes. In this way, the action can be incorporated into different narratives. On the other hand, the different dimensions of control can complement and reinforce each other. For example, rational control in conjunction with the deep commitments of the agent, exerts some pressure on volitional control in a snowball effect, so that a practical identity can be built up over time.

IV

In these last pages I will show how both the rational structure of planning agency and the way we interpret our own actions and those of others against this background, enable us to recognise that sometimes, when we attribute responsibility, we assume that other people are planning agents, or at least that the meaning of their behaviour makes sense against a temporally extended background. The intentions of agents, in conjunction with more complex structures, are part of narratives we develop about what happens, and in doing so, we impute certain events to others as part of what they are responsible for.

These narratives have various manifestations, I would like to draw attention to the relationship between two: scripts and schemata. AccordinG to Cristina Bicchieri:

> Schemata are cognitive structures that represent stored knowledge about people, events, and roles. When we apply a schema, our interpretation of the situation is theory-driven, in the sense that prior knowledge heavily influences the way we understand and interpret a salient stimulus. (2006: 93)

Bicchieri uses the expression “theory” because “a schema does not represent particular, detailed knowledge, but rather generic knowledge that holds across many instances”. Along this line, schemata have been represented as data structures that represent generic concepts and stereotyped situations stored in memory. In that sense, "schemata, unlike associations, are organic wholes comprised of parts that are oriented both to the whole and to other parts" (Casson, 1983: 431). Regarding its structure, "A schema generally includes a number of embedded subschemata as constituent parts, each of which interacts in its own right with elements in the environment." (Casson, 1983: 436). They are sequentially organized, that is, subschemata are embedded in a schema that is ordered to represent temporal or locational changes, cause-effect relationships, and the sequence of stages or actions in events. An example can show how this works:

> Eating in a restaurant, (…) is an event whose representation is comprised of a sequence of linked schemata-ENTERING, ORDERING, EATING, and EXITING. These schemata, termed "scenes," are in turn comprised of sequences of actions; TIPPING the waitress and PAYING the check are two of the constituent actions in the EXITING scene, for example. The sequence of actions defining scenes and larger-scale events is an elaborate causal chain: each action in the sequence results in conditions that enable the next action and must be completed before the next action can be started (CASSON, 1983: 438)

These sequences or chains are called scripts. Scripts connects specifics actions and states conditionally as the example shows. In Bicchieri’s words, scripts describe “a stylized, stereotyped sequence of actions that are appropriate in the context, and it defines actors and roles” (Bicchieri, 2006: 94). So, “When people use scripts, they know what to expect of each other. They need not be acquainted with each other or know of each other’s past performance. Their expectations are grounded in the certainty that, if indeed script s is being enacted, then actions a1, . . . , an will follow”. (2006: 96)

As we can see, scripts and schemata enable the interpretation of actions, CRAIGs and the attribution of intentions. When our friend gestures to ask for the bill after dinner and takes a credit card from her purse, we can say that she intends to pay the bill because the sequence of actions is the execution of a script that is part of the schema EXITING in the shared context. If we read her behaviour using the script, we are justified say that her behaviour has a purpose and is the implementation of a proximal intention. On the other hand, we can use schemata and scripts to develop our own intentions, whether they are proximal or distal: If we want to pay, we can stop her while she takes the wallet out of her pocket and so on. In summary, these cognitive structures help us read the actions of others and organise our own behaviour. If we know what to expect in certain situations, we can make sense of the actions of others and plan our movements at the same time. What is important is that these narratives sometimes assume that we are temporally extended agents.

To hold someone responsible it is necessary to attribute the wrong to a person. In order to do this, we use schemata and scripts and some of them reflect our temporally extended agency. In what follows, I will look at three of them. We can call them *temporal structures of imputation.*

The first structure is premeditation. If a person who wishes the death of another person studies the effects of a toxic substance on her and puts small amounts of that substance in her food until it causes her death, she may be considered to have committed premeditated murder. It is a temporally extended process, usually considered particularly serious because it involves the exercise of a large part of our rational faculties in order to achieve a wrong outcome and is also a clear expression of the deep commitments of the agent that are part of his origin control. None of these aspects of agency are present, for example, in accidental murder, and we usually react less intensely to the perpetrator in these cases.

The second structure is found in cases of negligence, where the origin of the risk can be traced to previous actions of the agent (v. Smith, 1983). An example of this is the case of a person who decides to wait to repair his car and who causes a traffic accident due to engine failure. In these

cases, although the proximate intention of the person does not mean that she intended the harmful outcome, it is possible to attribute that result to her prior action. In these cases, two points should be noted. The first is that what is being reproached is precisely the failure to exercise a temporally extended capacity to act. Since the agent creates the risk that causes the harm by a prior act of inadvertence, we are willing to regard her as culpable for the wrong outcome. Secondly, in this case, there is a counterfactual judgement that allows us to say what the agent should have done in the past to avoid what happened in the future, creating unity between the two temporally separate events.

Finally, there are the cases of *actio libera in causa*. In these cases, a person intentionally puts herself in a situation where she loses volitional control but does so with the aim of committing a crime under those circumstances (v. Hruschka, 1986). An example would be a person who, in order to commit a crime in a state of extreme drunkenness, ingests a large amount of alcohol beforehand, causing her to lose her mental faculties in committing the crime. In these cases, the criminal behaviour can plausibly be described as the circumstances known to the agent, and the overall situation planned by her, including her own loss of volitional control.

In these three types of cases, the control of the agent in the form of intentions and plans enables us to identify a common temporal structure which, in turn, leads us to impute the result to the agent by holding her responsible for it. This is true even though in cases of negligence there is no correspondence between the intention of the negligent act and the prior intention that grounds its occurrence, and that in cases of *actio libera in causa* the agent may lack volitional control while performing the criminal act.

If we do not assume that we are temporally extended agents, these structures of imputation would not have the meaning they have. In the first case, if we do not see the execution of the crime as the application of a premeditated plan, it is not possible to say why the perpetrator deserves a more intense punishment. In these cases, planning agency allows descriptions that include previous actions in what is considered contrary to the norm. In the other two cases, even if the norm-

violating behaviour does not satisfy the *mens rea* requirement, we understand that it is part of a plan or lack of a plan that should have been developed.

From these and other cases we can see that some structures of imputation are temporally constructed, as are the norms of imputation that regulates them (Figueroa Rubio, *forthcoming*). I believe that with this kind of framework we can explain why these structures make sense, showing on the one hand, what are the similarities and differences between them, and, on the other hand, how we understand our agency in its relation to responsibility.

References

Anscombe, G.E.M. 1963. *Intention*. second edition. Oxford: Blackwell

Bicchieri, C. 2006. *The Grammar of Society.* Cambridge: Cambridge University Press

Casson, R. 1983. “Schemata in Cognitive Anthropology” *Annual Review of Anthropology.* Vol. 12:429-462

Bratman, M. 1984. “Two Faces of Intention” *Philosophical Review* 93 (3):375-405

Bratman, M. 1999. *Faces of Intention.* Cambridge: Cambridge University Press.

Bratman, M. 2010. *Planning, Time, and Self-Governance.* Oxford: Oxford University Press.

Davidson, D. 1980. *Essays on Actions and Events.* New York: Clarendon Press.

Ferrero, L. 2017. “Intending, Acting, and Doing” *Philosophical Explorations* 20 (S2): S13–S39,

Figueroa Rubio, S. *forthcoming*. “Ascriptivism, Norms, and Negligence” in Shoemaker, D (ed) *Oxford Studies in Agency and Responsibility 8.* Oxford: Oxford University Press

Gallagher, S. 2020. *Action and Interaction.* Oxford: Oxford University Press.

Hruschka, J. 1986. "Imputation" *Brigham Young University Law Review.* 11: 669-710

Mele, A. 2019. “On snubbing proximal intentions” *Philosophical Studies* 176: 2833–2853

Moya, C. 2017. *El libre albedrío.* Madrid: Cátedra.

Smith, H. 1983. “Culpable Negligence” *Philosophical Review* 92 (4): 543-571

Sellars, W. 1980. “On Reasoning About Values” *American Philosophical Quarterly* 17: 81-101

Watson, G. (1996) “Two Faces of Responsibility”. Reprinted in Watson, G .2004. *Agency and Answerability*. Oxford: Oxford University Press.