

Ricerche

The Agent as Her Self: How Taking Agency Seriously Leads to Emergent Dualism

Maria Joana Rigato^{(α),(β)}

Ricevuto: 23 novembre 2015; accettato: 30 marzo 2016

Abstract To act is to be the author of an intentional bodily movement. I will show that, in order for that authorship to be assured, the agent must both amount to more than the mereological sum of her mental or neural states and events, and have an irreducible causal power over, at least, some of them. Hence, agent-causalism is the best position for any realist about action to assume. I will contend that, contrary to what many have claimed, agent-causalism is not an unscientific theory, since it can ground its view of the agent on a form of emergent dualism that can account for robust forms of agency without having to challenge the natural supervenience of the mental on the physical. I claim that the conditions of possibility for a causally effective emergent self are the presence of neuronal indeterminism and the break of causal closure, both of which will be shown to be compatible with our current scientific picture of the world.

KEYWORDS: Action; Self; Emergent Dualism; Downward Causation; Indeterminism

Riassunto *Prendere sul serio l'agente: ovvero, come un'attenta analisi dell'agentività ci conduce al dualismo emergentista* – Agire vuol dire essere l'autore di un movimento corporeo intenzionale. Intendo mostrare che, per assicurarsi il riconoscimento di essere l'autore di un'azione, l'agente debba essere più della semplice somma mereologica dei propri stati mentali ed eventi neurali e che debba anche avere un irriducibile potere causale su di essi, o al limite su parte di essi. Pertanto, una concezione causale riferita all'agente è la migliore posizione da assumere per chi voglia dirsi realista riguardo alle azioni. Intendo affermare che, diversamente a quanto da molti sostenuto, la concezione causale riferita all'agente non sia una teoria non scientifica, dal momento che può fondare la sua concezione dell'agente su una forma di dualismo emergentista in grado di legittimare forme robuste di agentività senza dover mettere in dubbio la sopravvenienza del mentale sul fisico. Indeterminismo neurale e rottura della chiusura causale saranno indicate come condizioni di possibilità per un sé causalmente ed effettivamente emergente e mostrerò come entrambi siano compatibili con l'immagine scientifica del mondo che oggi abbiamo.

PAROLE CHIAVE: Azione; Sé; Dualismo emergentista; Causalità rivolta verso il basso; Indeterminismo



Introduction

AS A PATIENT SUFFERING FROM Anarchic

Hand Syndrome once stated: «I'd light a cigarette, balance it on an ashtray, and then my left hand would reach forward and stub it

^(α)Centro de Filosofia das Ciências, Universidade de Lisboa (CFCUL), Campo Grande, Edifício C4 – 1749-016 Lisboa (PT)

^(β)Fundação Champalimaud, Avenida Brasília – 1400-038 Lisboa (PT)

E-mail: joana.rigato@neuro.fchampalimaud.org (✉)



out. It would take things out of my handbag and I wouldn't realize it, so I would walk away. I lost a lot of things before I realized what was going on».¹

What is it that allows anyone reading this paper to distinguish his/her ability to act from the behavior of the alien limb depicted in this quotation? Is it the fact that we are conscious of what we do? Is it that we act according to our will?

In philosophy of action the most popular view about voluntary action is an event-causal view, according to which actions are events caused in an appropriate way by other events, namely, the agent's having certain beliefs and desires.² More often than not, event-causalism about actions comes along with a reductionistic account of the agent, considered as the epistemologically emergent but ultimately reducible collective entity whose causal powers are explainable (if not in practice, at least in principle) by the causal powers of its parts.

My thesis is that event-causal and reductive accounts are incapable of providing an adequate justification for the distinction people intuitively and practically make between actions and sub-actional behaviors. Such accounts should be replaced by an agent-causalist view, which is explanatorily more powerful. In order to show why this is so, I find it useful to start with some empirical considerations.

■ Actions and agents

In neuroscience, action is defined in opposition to response.³ There is a continuum in animal behavior at the farthest end of which we find simple reflexes (immediate and automatic motor responses), while at the other extreme end lie voluntary actions (not directly determined by any external stimulus).

Animal agents (human or not) enjoy a sort of «freedom from immediacy»⁴ by which they are able to make a decision or to self-initiate a spontaneous movement in the absence of clues from the environment that

might serve as secure evidence in favor of that choice over any alternative.

The agent's recognition of herself as an entity distinct from the environment and from the others, and her identification with that part of the physical world that she can control to some degree – her body – is a precondition for any voluntary action to take place, as well as for it to be interpreted as such by others. In the words of neuroscientist Björn Brembs, «in order to understand actions, it is necessary to introduce the term self».⁵

An animal can distinguish its self from the world via a mechanism called refference,⁶ whereby it can naturally and unconsciously tell apart those sensory stimuli that are consequences of its own actions and thereby are under its control (e.g. the darkness caused by its eye blinks), from those that are not. Even though it is reasonable to assume that not all animals are capable of controlling what they do with their bodies, the fact that even simple invertebrates can distinguish their spontaneously generated movements from their elicited responses to the environment opens up the possibility for both agency and selfhood to evolve in the animal kingdom. When a certain degree of complexity in the biological hierarchy of species is reached, a sort of owner/body distinction⁷ arises which allows for the animal to intentionally execute its power of self-movement.

To sum up, just as only a subset of events can be considered behaviors (the arrival of a photon at my retina is not a behavior of mine; my eyes blinking is), only a subset of behaviors can be classified as actions. Actions are behaviors that are performed intentionally, that is, behaviors in which the agent was the “author” of the bodily movement. But what does this authorship amount to? That is what the whole dispute in philosophy of action is about.

Some clarification of terms is in order. For an act to be intentional is for it to be the execution of a plan, which is the mental representation of the future action to be performed. The intention is an executive atti-

tude towards that plan, a disposition to put it into practice.⁸ Previous to the conception of that plan, of course, the agent must have certain reasons to do something or to act in such a way as to achieve a certain goal.

Reasons are usually taken to be pairs of desires and beliefs. Thus, the desire to drink plus the belief that water is in the fridge and that the enactment of certain behaviors is necessary if I am eventually to swallow it, will lead me to form a certain plan and eventually to execute it, opening the fridge, taking out the water, filling up a glass with it and drinking it.

It is very important to note, however, that there are many purposive behaviors that are apparently intentional and caused by the agent's reasons, but which cannot be considered to be actions because the production of the intention that brought them about was not controlled by the agent. Let us use the example of addiction. The difference between the behavior of drug addicts and non-addicts is not that the former is not driven by reasons. Addicted behavior is, in fact, motivated by certain desires (e.g. the desire for the drug) and certain beliefs (e.g. the belief that shooting up a certain dose in a certain manner will provide the desired effect). However, in cases of deep addiction there is no intermediate agential intervention between the reasons and the intention, nor between the intention and the behavior. To use Harry Frankfurt's words, an addict's

desire to take the drug will be effective regardless of whether or not he wants this desire to constitute his will.⁹

The drug addict goes about his business in autopilot mode and this is why his behavior cannot yet be considered an action. The fact that it was not the agent's self that chose to act as she did, but it was rather a sequence of blind events in her brain that brought that behavior about, renders these cases similar to the obsessive-compulsive response to urges in cases of Obsessive Compulsive Disorder¹⁰ or in other neurological diseases such as Tou-

rette's Syndrome.¹¹

Touretters, in fact, can help us make even clearer the need for a distinction between actions and sub-actional behaviors, based on the intervention of the agent's will as the action's ultimate source. A patient suffering from Tourette Syndrome is always self-aware and lives a fruitful life, with a job, a family and friends. However, her condition makes it very difficult (often impossible) for her to control the outburst of motor and vocal tics, some quite simple such as barking or eye blinking, others much more complex, such as punching herself, touching objects or people, bending and twitching her body, uttering inappropriate sentences (coprolalia) or repeating what other people say (echolalia). The disease makes her also more prone to many other behavioral symptoms such as obsessive thoughts, negative reactions to novel situations due to anxiety, great difficulty in inhibiting impulsive behavior, episodes of uncontrollable rage, etc.

When a patient's tics are involuntary and uncontrollable, her behavior is merely reactive. In those cases, tics just have to be released, like a sneeze. However, while growing up, most Touretters start experiencing premonitory sensory phenomena which might allow them to sense that a certain tic is about to arise and to prevent it occasionally, due to some training. This is not easy: even when it is possible to prevent the tics, that prevention costs the patient a lot of effort, it increases stress and it can only last for a short period of time. Nevertheless, the degree of control a patient can have over her body, albeit limited, may allow her to avoid the social disadvantages of the tic, like being stared at and misjudged by other people. Let me present two examples of such a tension between the voluntary inhibitory breaks versus the involuntary outburst of the tic:

One woman, growing up on a farm, took several long walks daily in the woods. Her family attributed these to a solitary or soulful nature; in reality, she told me, she simply needed a place where she could re-

lease her tics, which she had to suppress in the house. She would walk for miles, “twitching and spitting like a maniac”, then return home unsuspected.¹²

In a situation of close social contact, where satisfying the urge for a facial or neck tic would be very noticeable, almost all of my informants said that they would occasionally induce a tic in the leg or foot instead. By performing that tic intensively – clenching the toes hard or hyperextending the ankle, for example – they could divert energy away from the facial tic, and perhaps suppress it altogether.¹³

The possibility of intentionally blocking the tic for some time opens the door for the behavior of a Touretteur to be much more voluntary than what *prima facie* might appear. When the patient blocks her tics, she becomes the helmswoman of her ship – her body – steering it according to her decisions, in spite of the highly conditioned elbow room at her disposal. Like Timothy Schroeder put it:

The way in which a Tourettic individual resists an urge to tic says much about the quality of the individual’s will, but the urge itself says nothing.¹⁴

It is important to note, however, that when the TS patient gives in to the tic, she feels that she is actually “doing” it, that she is not assisting passively at the event of the tic coming to be. The urge to tic is a conscious mental state akin to other more common desires (impulses, needs) which becomes the direct cause of the Touretteur’s behavior when she tics. I consider the urge to tic to be a reason to act, just like rage can be a reason to act: they are both mental states that can motivate the agent directly, when she renounces to exerting control over her behavior via an actively formed intention.¹⁵ When hot-tempered people are furious and engage in a discussion or a fight, they often say and do things they might regret and that they had

explicitly promised themselves not to say or do. These are common situations in which it is not a drink or a drug that take hold of our reactions, it is not a disease either, it is our emotions. In certain situations, they make us lose control over our behavior and sometimes say “it wasn’t me, it was my resentment [or rage, or fear or jealousy] speaking”.¹⁶

Cases of substance abuse, psychiatric disorders, neurologic diseases and emotionally driven behavior help us see how agency cannot be reduced neither to neural states and events, nor to their mental correlates. When the agent acts, she does not do so merely by “having” certain beliefs and desires; she “activates” those reasons, and connects them to her intentions to act. The agential role in action can be described as: (1) forming an intention to act for certain reasons and (2) producing a bodily movement according to that intention. It is not the reasons that produce the intention *per se*, nor the intention that produces the movement. If a mere causal link between these elements were sufficient, then we would be unable to justify why, under some circumstances (which might be abnormal, like the neural pathologies just mentioned, as well as quite regular, as in cases of emotionally driven behavior), the agent can fail to participate in the behavior that she is supposed to be the author of. After all, in all these cases, reasons and even intentions (e.g. the intention to hurt my opponent in a fight) are present just the same.

But should one not try to find a way out of this problem that is still event-causal and reductionistic, before giving in to a substance-causal or dualistic alternative? David Velleman famously argued for such a solution, proposing that the involvement of the agent in action be identified with her desire to act for reasons.

The agent, in his capacity as agent, is that party who is always behind, and never in front of, the lens of critical reflection, no matter where in the hierarchy of motives it turns. What mental event or state might

play this role of always directing but never undergoing such scrutiny? It can only be a motive that drives practical thought itself [...] What animates practical thought is a concern for acting in accordance with reasons. And I suggest that we think of this concern as embodied in a desire that drives practical thought.¹⁷

Can we follow Velleman in this reduction of the causal role of the agent to the desire to act for reasons, the desire to do what makes sense and is intelligible to her?

I believe that the main problem with this proposal is that the desire to act for reasons is too narrow. An action can be authored by the agent regardless of how the agent can herself explain it. It is entirely coherent to conceive the experience of doing things that we do not understand but which we fully endorse as actions that we performed intentionally: many of us (if not all) have had this experience some time or another. Also, we often act akratically, acting from reasons that we do not consider to be our best, and this does not prevent us from being accountable, nor others from giving us credit or blame for our acts – which reveals how much we are considered to have actively contributed to the action's coming to be.¹⁸

But are these counterexamples a problem related to the specific attitude that Velleman chose as being functionally identical to the agent (the desire to act for reasons), or are they representative of a more general drawback capable of affecting any similar type of account?

I believe that no reduction of the agent to specific mental states of hers can appropriately respond to the problem of how to account for the agential authorship of actions. Velleman himself claims:

What makes us agents rather than mere subjects of behaviour – in our conception of ourselves, at least, if not in reality – is our perceived capacity to interpose ourselves into the course of events in such a way that the behavioural outcome is traceable directly to us.¹⁹

I agree entirely. But I suspect that, no matter what psychological states and events a reductionist might elect as the core elements that can «speak for the agent»,²⁰ there will always be an available counterexample of an action that lacks that element in its etiology but which we are willing to count as an action nonetheless. What is undoubtedly present in all actions in such a way that we «interpose ourselves into the course of events» is not a state but rather an ability: it is the agent's power to form an intention to act and, through that intention, to be the cause of the action. It is this ability – which I call the agent's will – that an agent can fail to exercise in cases of addiction, for example, and it is this ability that will now lead me to postulate the existence of the agent as part of the total cause of an action.

■ The merits of the agent-causalist stance

The fact that the agent's ability to form intentions to act is irreducible to her mental states does not imply the negation of the causal power of her reasons. The agent's beliefs and desires do influence her decision, of course, but there would be no action were the agent not to actively determine which reasons will be effective.

Therefore, I contend that agency as such presupposes the presence of an agent-cause who can downwardly influence the weight her reasons and their neural correlates will have in the sequence of events taking place in her brain, her body, and the external world. That agent-cause (we may call it the agent's "self") must be an irreducible entity, otherwise her causing would reduce to the micro-causings of her parts (her mental states and events or their neural correlates).

If the agent were nothing more than the mereological sum of her mental states and events and their neural correlates (a «bundle or collection of different perceptions»,²¹ quoting Hume) and her intending could be reduced to her intention being brought about by some of those, then there would be noth-

ing, besides them, which might influence the behavioral outcome.

A “humean” agent is a composite entity with structural properties,²² just like rocks or plants are, and thus her bringing it about that she will do A instead of B is actually the result of each of her parts’ causing a certain complex collection of events at the mental as well as the neural level (independently of which theory one may have about the relation between the two). Thus for an action to be brought about by the agent, as opposed to her being passive relative to the occurrences taking place within her, the agent as such must be the cause of the action, by willing.

My account of action, then, has two major metaphysical implications:

- 1) that the agent’s self must be more than the mereological sum of her mental states and events
- 2) that she must have a sort of downward causal power over the goings-on in her mind and brain, rather than being the mere locus where the causal chain from reasons to intentions and to actions takes place

Both these postulates, as well as the main claim that actions are caused directly by agents, are endorsed by agent-causalism, a position that has been mistakenly associated with libertarianism²³ in the context of the free will debate. I claim this has been a mistake because compatibilists about free will (that is, people who consider that an action can be free even if it was deterministically caused) should care about the irreducibility of the agent’s self just as much as their opponents.²⁴

Both compatibilists and incompatibilists agree that self-determined action entails physical and moral responsibility insofar as it involves control,²⁵ and control is another word for authorship. If the agent were not the author of the action, her behavior would be something that “happens” to her and not something that she “does”. Like we saw in the previous section, the agent’s power lies in the

fact that she herself (and not the psychological and neural events that happen within her) is the action’s author. Without the controlling agent, there are no true actions in the world, regardless of its causal nature.

Therefore, agent-causalism is the position that any realist about action should assume. However, it has been discredited as an anti-scientific position. In the remainder of this paper, I will contend that this is wrong, for the claims of agent-causalism do not contradict our current best science.

■ Emergent dualism and a case for the plausibility of its conditions of possibility

Advances in neuroscience keep corroborating the thesis that the mind supervenes²⁶ on the brain: every mental change is anchored on a physical change and every mental state corresponds to a neural correlate, the manipulation of which leads to mental consequences. Psychiatric drugs are good examples of this, as well as our scientific practice and our daily experience of figuring out how the world works. If we put two glasses of milk in the refrigerator at the same time and one of them gets spoiled sooner than the other, we will first look for a visible difference between them: distinct packs with different expiring dates or some dirt in one of the glasses, for example.

If we cannot find it, we will likely think there is an invisible reason that can explain what happened and explain also why it happened in only one of the glasses rather than in both. We will intuitively postulate a microscopic difference which will surely have caused the macro phenomenon that we can detect with our senses. In the same way, when scientists encounter differences between two samples of a similar substance at the biological level that they cannot easily explain, they will look for corresponding differences at the chemical level that might justify the phenomenon.

This is how science reasons and how it has progressed. Differences at one level lead researchers to look for differences at the next

smaller one where explanations might be found, and along that trail they discover the patterns of the microscopic structure of reality, like the periodic table or the standard model of particle physics, and find out the genetic or viral origin of certain diseases. Therefore, I believe that to question supervenience would come at a very high cost. However, unlike what is commonly assumed, agent-causalism does not have to challenge it.

I contend that emergent dualism²⁷ is the type of account that can best integrate the intuitive view that the agent is the author of her actions and the highly plausible supervenience thesis according to which there is no autonomous mind capable of wandering off independently from the goings-on taking place in the brain. Emergent dualism states that the self is distinct from its body but emerges from it naturally and depends on it throughout life.

This type of account relies on the concept of ontological emergence. Ontological emergence is a relation between different-level entities, in which the upper-level entity depends on the lower-level structure but possesses new causal powers that cannot be explained only on the basis of the properties below. These new causal powers are often taken to have downward causal effects over the lower-level substrate that brings them about, which seems quite puzzling to reductionists. In the case of an emergent entity such as the self, these causal powers may be, for example, the power to experience interactions with the world from a first-person perspective (i.e. phenomenal consciousness) and the causal power this paper is mostly concerned with: that of actively forming intentions to act (i.e. the agent's will).

I contend that ontological emergence does not imply the break of supervenience as it requires only two conditions of possibility, both of which are as scientifically plausible as their rival theses:

1) that the emergence base works indeterministically

2) that the physical world is not causally closed.

I will now address them both.

■ The requirement of bottom-level indeterminism

The fact that there must be indeterminism²⁸ at the bottom-level in order for new causal powers to emerge and to have downward effects is very seldom pointed out by emergentists, which is something I find very surprising. To use the example of an action, which is what concerns me here, imagine the case when I purposefully wave at a friend that I see across the street. For my intention to effectively cause my behavior, it must be the case that some particles in my motor cortex have different possibilities of movement at the instant immediately following my decision to move, even given all the specifications of their circumstance at the instant of my decision (the precise values of their mass, charge, location, etc. as well as the complete state of each of their neighbor particles); only if this is so, can they initiate a causal chain that will lead all the other particles correlated with them, namely the ones which constitute my hand, to move according to my will.

There must be diverse "possibilia" in my brain and body that may or may not become "actualia" by downward constraint, otherwise there would be nothing left for the emergent self to cause. Its causal power would be redundant and therefore epiphenomenal.

Many have defended that the question whether the world works deterministically or not and the question whether emergence is possible are two independent problems. Arguing for the effectiveness of mental causation, Roger Sperry²⁹ has famously coined the example of a wheel rolling downhill in which each molecule's movement is determined in space and time by the overall properties and dynamics of the wheel as a whole. Even though Sperry's interpretation of such a case

as an example of emergence and its use for an analogy with the mutual interdependence between consciousness (the rolling wheel) and the brain (whose individual neurons are «carried along» just like the molecules in the wheel) have both been generally criticized in the literature, I believe this example is useful for the point I wish to make.

Sperry defended micro and macro determinism in the neural substrate and functioning of the brain, and also argued that, in exerting their «supervenient downward control», the emergent mental properties could not intervene nor disrupt the causal activity at the lower-level.³⁰ Given these two contentions, the analogy he used for the reciprocal interaction and determination between mental and neural levels *had to be* that of the trivial but only apparent mutual influences between the micromolecular level of the atoms in the rolling wheel (which are obviously causally efficacious at their own level), and the determination of their space-time trajectories by the entity as a whole (which is actually just a macro level of description that does not correspond to any real superior and novel causal influence over the wheel's components).

I believe Sperry's use of the example of the wheel is particularly significant in the context of his explicit defense of the compatibility of emergence and universal determinism, insofar as it shows precisely how the Newtonian laws that govern or describe the behavior of physical entities do not leave any room for downward causation. What does it mean for atoms and molecules to be «carried along» as the wheel rolls downhill? The rolling of the wheel itself is nothing but the sum of the movements of its atoms and molecules, which lower level laws manage perfectly to describe within the framework of a reductionist physics. But maybe the example of the wheel was just unfortunate. Could a better instance of deterministic emergence help?

Philosopher and physicist Robert Bishop often uses Rayleigh-Bénard convection cells as a paradigmatic example of complex systems in which sensitive dependence allows

for the emergence of higher level structures with downward causal power – in other words, a “control hierarchy”. However, even though an analogy between this type of system and the human brain would be more accurate than the preceding example because of its increased complexity, still the movement of each molecule in a fluid convection cell is constrained by the movement of all the other elements of the fluid (in a horizontal all-to-one sort of constraint). The whole of the system as such is more than the mere sum of its parts in the sense that the dynamics itself must also be accounted for by any faithful model of the system, but its causal power cannot be manifested over and above the causal power of each one of its parts if the trajectories in space-time that they follow are deterministic. Unless the behavior of the components obeys probabilistic laws that endow it with alternative futures, the dynamics of the system as a whole is only an emergent epiphenomenon and the interaction among particles is what really calls the shots.

■ The plausibility of neuronal indeterminism

Is any of what has just been stated a problem for the substance-causalist? Not at all. First, because even if we may not be able to know whether our neural activity leaves room for alternative possibilities, there is no scientific evidence to this day that the human brain works deterministically either. Second, because we have increasing scientific evidence that quantum events can have macroscopic effects.

What neuroscientists deal with on a daily basis are stochastic processes. The question whether these processes are only indeterministic at an epistemic level or are actually the macro manifestation of more fundamental indeterminacies is something empirical research cannot tell us just as yet. Adina Roskies, philosopher and neuroscientist, explains this very clearly:

The picture that neuroscience has yielded

so far is one of mechanisms infused with indeterministic or stochastic (random or probabilistic) processes. Whether or not a neuron will fire, what pattern of action potentials it generates, or how many synaptic vesicles are released have all been characterized as stochastic phenomena in our current best models. However, whether the unpredictability we perceive is really due to fundamentally indeterministic processes, or to complex deterministic ones beyond our present understanding is something neuroscience cannot tell us.³¹

The problem with expecting science to provide an answer to the question of determinism is that it is actually a metaphysical question, not an empirical one. Our quest for the causes of phenomena deep into ever more fundamental layers of reality is never concluded: one can always postulate one more level underlying the ones we have come to know well. That is why Roskies says that neuroscience will “never” be able to give us a definitive answer to the mystery of neuronal indeterminism:

Because a deterministic system can radically diverge in its behavior depending on infinitesimal changes in initial conditions, no evidence for indeterminism at the level of neurons or regions of activation will have any bearing on the fundamental question of whether or not the universe is deterministic. That is ultimately a question for physical theory, and will be answered by our best theory of the fundamental nature of physics, not at the level of brain science.³²

So even if neuroscientists deal with epistemic indeterminism at the neural level, the nature of the causal interactions taking place underneath is out of their reach. And maybe the causal nature of the world will always remain ultimately inaccessible to the observer, even at the microphysical level of analysis.

Nevertheless, the fact that what we know

so far meets exactly what would be expected “if” the brain worked indeterministically deserves to be acknowledged. Neuroscience has not discredited the hypothesis of neurological indeterminism and it keeps accumulating evidence that is entirely consistent with it.

Moreover, the fact that this question has not been settled yet has not prevented major neuroscientists such as William Newsome³³ or Paul Glimcher,³⁴ along with many others,³⁵ from expressing their conviction in favor of an indeterministic account of the nature of brain processes. From their point of view, the indeterminacy that we find at the behavioral level is a result of genuinely random events at the cellular and subcellular levels, like for instance the patterns of vesicular release and the variations in membrane voltage, which seem to be «the product of interactions at the atomic level, many of which are governed by quantum physics and thus are truly indeterminate events».³⁶

In the article *Indeterminacy in Brain and Behavior*, in which he presents a long review of the neuroscientific literature related to the problem of the source of variability in the brain, Glimcher concludes: «Physical indeterminacy seems to be a fundamental property of the brain».³⁷

Of course, this depends on the general possibility of quantum fluctuations having macroscopic effects in warm and wet environments such as the brain. Some years ago this might have seemed much harder. Phenomena like superconductivity or the Bose-Einstein condensates were usually cited as examples of indeterministic effects at the macro scale but they required temperatures close to absolute zero. Today, however, there is increasing evidence that functional quantum effects operate in biology as well,³⁸ with cases stemming from photosynthesis to bird brain navigation, processes in which the effects of quantum phenomena are amplified in warm and wet systems and have chemical consequences.³⁹

More than ever, the argument that macro objects such as the brain must function deterministically and that quantum phenomena

can never be amplified enough to have effects at a level that is relevant for action production seems unfounded.

■ The break of Causal Closure

If the causes at the bottom-level are insufficient to determine the full development of the biological system we call the agent's body, then the gaps that they leave open can be filled by the extra causes provided by the emergent level. This of course depends on the break of a principle that very often remains unquestioned: the Causal Closure of the Physical (CCP). But so much the worse for this principle, then, which can never be proven nor disproven by science and which is merely a premise in most of the arguments in favor of physicalism!⁴⁰ As Robert Bishop put it:

Physics itself does not imply its own causal closure nor is there any proof within physics of its own completeness.⁴¹

Evidence from physics supports only a qualified reading of the Causal Closure principle as a typicality condition (stating what happens in scientific labs, under controlled circumstances that prevent non-physical interferences). According to this reading, championed by Bishop among others, what CCP tells us is that "in the absence of non-physical influences", physical causes (events and laws) will produce physical effects.

In order for Causal Closure to entail the ineffectiveness of non-physical causes in the etiology of physical effects, one would have to assume also «that the only efficacious states and causes are physical ones»,⁴² which is a postulate we have no evidence for and which begs the question of physicalism.⁴³

Hence, the reductionist physicalist is left with two unattractive alternatives: on the one hand, she can endorse a strong but unjustified interpretation of CCP, which would prevent any emergent entities from having autonomous downward causal powers (preventing also any intentional action from tak-

ing place, according to the argument developed in the first two sections of this paper), but which cannot be confirmed nor disproven by empirical means. On the other, she can adopt the weaker typicality version, supported by physical science, but which is insufficient to ensure that only physical causes are effective. Either way seems to get reductionism into more trouble than anticipated.

■ Conclusion

Let us now sum up the main steps of my argument. Agency, as opposed to mere subactional behavior, requires the intervention of an agent-cause in the causal chain leading from reasons to action. Agent-causation, in turn, implies the existence of an irreducible self with downward causal powers over the neural and mental states and events taking place within the agent's body. The plausibility of neuronal indeterminism, together with the fragility of the principle of causal closure, allow for the postulation of the ontological emergence of the agent's causal powers, such that natural supervenience is saved but there is genuine novelty at the mental level, with the possibility of downward constraint.

Unlike the paramecium, which moves reactively only, animal agents are capable of directing their bodily movements and interactions with the world purposefully. And this faculty is based on the owner/body distinction whereby their emergent self can control the movements it chooses to engage in.

In the title of this paper I claimed that agent-causation leads to emergent dualism. I hope to have shown why this is so. I hope also to have provided good reasons in favor of both these positions (agent-causalism, as well as emergent dualism), which are serious and plausible alternatives to the mainstream reductionist views of the mind, the self and agential control.

■ Notes

¹ Karen Byrne, a patient suffering from Anarchic

Hand Syndrome after a surgery to cure epilepsy in which her brain's *corpus callosum* was cut (article by Dr. Michael Mosley for *BBC News Health* 01.20.2011).

² This is what David Velleman called the «standard story» (J.D. VELLEMAN, *What Happens When Someone Acts?*, in: «Mind», vol. CI, n. 403, 1992, pp. 461-481), which is often credited as the “davidsonian” view of action, after Donald Davidson’s influential (D. DAVIDSON, *Essays on Action and Events*, Clarendon Press, Oxford 1980).

³ Cf. B. BREMBS, *Toward a Scientific Concept of Free Will as a Biological Trait: Spontaneous Action and Decision-making in Invertebrates*, in: «Proceedings of the Royal Society B: Biological Science», vol. CCLXXVIII, 2011, pp. 930-939; P. HAGGARD, *Human Volition: Towards a Neuroscience of Will*, in: «Nature Reviews Neuroscience», vol. IX, 2008, pp. 934-946.

⁴ Cf. M.N. SHADLEN, J.I. GOLD, *The Neurophysiology of Decision-making as a Window on Cognition*, in: M.S. GAZZANIGA (ed.), *The Cognitive Neurosciences*, MIT Press, Cambridge (MA) 2004, III ed., pp. 1229-1241.

⁵ B. BREMBS, *Toward a Scientific Concept of Free Will as a Biological Trait*, cit., p. 936.

⁶ E. VON HOLST, H. MITTELSTAEDT, *Das Reafferenzprinzip. Wechselwirkungen zwischen Zentralnervensystem und Peripherie*, in: «Naturwissenschaften», vol. XXXVII, n. 20, 1950, pp. 464-476.

⁷ Cf. Helen Steward, who has recently proposed an interesting definition of the agent as «an entity that has a body and can make that body move in various ways» (cf. H. STEWARD, *A Metaphysics for Freedom*, Oxford University Press, Oxford 2012, p. 32).

⁸ Cf. A. MELE, *Effective Intentions. The Power of Conscious Will*, Oxford University Press, New York 2009, pp. 3-7.

⁹ H.G. FRANKFURT, *Freedom of the Will and the Concept of a Person*, in: «Journal of Philosophy», vol. LXVIII, n. 1, 1971, pp. 5-20, here p. 20.

¹⁰ Cf. S.E. SWEDO, J.L. RAPOPORT, H. LEONARD, M. LENANE, D. CHESLOW, *Obsessive-Compulsive Disorder in Children and Adolescents. Clinical Phenomenology of 70 Consecutive Cases*, in: «Archives of General Psychiatry», vol. XLVI, n. 4, 1989, pp. 335-341.

¹¹ Cf. T. BANASCHEWSKI, W. WOERNER, A. ROTHENBERGER, *Premonitory Sensory Phenomena and Suppressibility of Tics in Tourette Syndrome: Developmental Aspects in Children and Adolescents*, in: «Developmental Medicine and Child

Neurology», vol. XLV, n. 10, 2003, pp. 700-703; A. BUCKSER, *Before Your Very Eyes: Illness, Agency, and the Management of Tourette Syndrome*, in: «Medical Anthropology Quarterly», vol. XXII, n. 2, 2008, pp. 167-192.

¹² A. BUCKSER, *Before Your Very Eyes*, cit., p. 175.

¹³ *Ivi*, p. 176.

¹⁴ T. SCHROEDER, *Moral Responsibility and Tourette Syndrome*, in: «Philosophy and Phenomenological Research», vol. LXXI, n. 1, 2005, pp. 106-123, here p. 119.

¹⁵ Cf. R. HOLTON, K. BERRIDGE, *Addiction between Compulsion and Choice*, in: N. LEVY (ed.), *Addiction and Self-Control: Perspectives from Philosophy, Psychology, and Neuroscience*, Oxford University Press, New York 2013, pp. 239-268.

¹⁶ A similar example is suggested by Velleman in his *What Happens When Someone Acts?*, cit., p. 465.

¹⁷ *Ivi*, pp. 477-478.

¹⁸ Cf. Nomy Arpaly and Timothy Schroeder’s critique to Velleman’s proposal on the basis of examples of inverse akrasia: N. ARPALY, T. SCHROEDER, *Praise, Blame and the Whole Self*, in: «Philosophical Studies», vol. XCIII, 1999, pp. 161-188.

¹⁹ J.D. VELLEMAN, *What Happens When Someone Acts?*, cit., p. 465.

²⁰ M.E. BRATMAN, *The Structure of Agency: Essays*, Oxford University Press, New York 2007, p. 4. Bratman too proposed a reductionist account of self-determination, as a solution to what he calls the problem of «agential authority».

²¹ D. HUME, *A Treatise of Human Nature* (1738), in: *A Treatise of Human Nature by David Hume. Reprinted from the original edition in three volumes, Book I*, edited by L.A. SELBY-BIGGE, Clarendon Press, Oxford 1960, p. 252.

²² Cf. O’Connor’s definition of structural properties: «A property, S, is structural if and only if proper parts of particulars having S have some property or properties not identical with S, and this state of affairs is constitutive of the state of affairs of the particular’s having S» (T. O’CONNOR, *Persons and Causes*, Oxford University Press, New York 2000, p. 109).

²³ Libertarianism is the view according to which: (1) free will is impossible in a deterministic world; (2) people are sometimes free; (3) this means that our world must work indeterministically, at least to some degree. Two of the most renowned contemporary agent-causal libertarians are Timothy O’Connor (T. O’CONNOR, *Persons and Causes*, cit.) and Jonathan Lowe (J. LOWE, *Personal Agen-*

cy, Oxford University Press, New York 2008).

²⁴ Cf. C.E. FRANKLIN, *If Anyone Should Be an Agent-Causalist, then Everyone Should Be an Agent-Causalist*, in: «Mind», forthcoming. See also N. MARKOSIAN, *A Compatibilist Version of the Theory of Agent Causation*, in: «Pacific Philosophical Quarterly», vol. LXXX, n. 3, 1999, pp. 257-277.

²⁵ What compatibilists and incompatibilists disagree about is what this control amounts to. Cf. J.M. FISCHER, M. RAVIZZA, *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge University Press, Cambridge 2000.

²⁶ I am defending the high plausibility of “natural supervenience” of non-fundamental entities on the bottom-most metaphysical domain, intended as the asymmetric covariance relation by which all macro changes that those entities undergo depend on corresponding micro changes, but not vice-versa. Note that this is distinct from the relation of “logical supervenience” which holds between properties A and B such that there cannot be a world in which A (the subvenient property) is present without B (the supervenient property) being present as well. This stronger form of supervenience (which is the type of relation one can argue does not hold between the brain and consciousness, thus allowing for the conceivability of zombies), is not the relation I am arguing one should assume exists between mental states and their neural correlates (David Chalmers presented this distinction clearly in D. CHALMER, *The Conscious Mind: in Search of a Fundamental Theory*, Oxford University Press, New York 1996).

²⁷ Cf. W. HASKER, *The Emergent Self*, Cornell University Press, Ithaca 1999.

²⁸ I take determinism to be the causal nature of a world in which, given a full description of all its elements and laws at t_1 , only one possible state can follow at t_2 . Indeterminism is the negation of determinism.

²⁹ Cf. R.W. SPERRY, *A Modified Concept of Consciousness*, in: «Psychological Review», vol. LXXVI, n. 6, 1969, pp. 532-536.

³⁰ These two theses, so I argue, are interdependent, thus I believe Sperry's contention of the truth of the latter (that emergent laws do not contradict the lower-level ones) was actually a coherent consequence of his adopting the former (determinism).

³¹ A. ROSKIES, *Neuroscientific Challenges to Free Will and Responsibility*, in: «Trends in Cognitive Sciences», vol. X, n. 9, 2006, pp. 419-423, here p. 420.

³² *Ivi*, pp. 420-421.

³³ W. NEWSOME, *Neuroscience, Explanation and*

the Problem of Free Will, in: W. SINNOTT-ARMSTRONG (ed.), *Moral Psychology. Vol. IV: Free Will and Moral Responsibility*, MIT Press, Cambridge (MA) 2014, pp. 81-95.

³⁴ P. GLIMCHER, *Indeterminacy in Brain and Behavior*, in: «Annual Review of Psychology», vol. LVI, 2005, pp. 25-56.

³⁵ Most notably, Björn Brembs has been focusing part of his neuroscientific research on the intrinsic source of behavioral variability in biological organisms, a topic on which he has a clear position in favor of neural sensitive dependence on quantum fluctuations (B. BREMBS, *Toward a Scientific Concept of Free Will*, cit.).

³⁶ P. GLIMCHER, *Indeterminacy in Brain and Behavior*, cit., p. 49.

³⁷ *Ibidem*.

³⁸ In a recent article, Stuart Hameroff and Nobel laureate Roger Penrose, probably the two most famous defenders of the idea that consciousness is a product of quantum effects in the brain, cited many studies which purport to show evidence of quantum effects in biological processes such as: ion channels, sense of smell, DNA, protein folding, and biological water (S. HAMEROFF, R. PENROSE, *Consciousness in the Universe. A Review of the 'Orch OR' Theory*, in: «Physics of Life Reviews», vol. XI, n. 1, 2014, pp. 39-78, here p. 63).

³⁹ There is in fact a new field of research called Quantum Biology, dedicated to studying non-trivial quantum phenomena in biological systems (Cf. S.F. HUELGA, M.B. PLENIO, *Vibrations, Quanta and Biology*, in: «Contemporary Physics», vol. LIV, n. 4, 2013, pp. 181-207). About electronic quantum effects which occur at ambient temperatures in proteins involved in photosynthesis see G.S. ENGEL, T.R. CALHOUN, E.L. READ, T.-K. AHN, T. MANCAL, Y.-C. CHENG, Y.-C., R.E. BLANKENSHIP, G.R. FLEMING, *Evidence for Wavelike Energy Transfer Through Quantum Coherence in Photosynthetic Systems*, in: «Nature», vol. CDLXVI, n. 7137, 2007, pp. 782-786; H. LEE, Y.C. CHENG, G.R. FLEMING, *Coherence Dynamics in Photosynthesis: Protein Protection of Excitonic Coherence*, in: «Science», vol. CCCXVI, n. 5830, 2007, pp. 1462-1465; I.P. MERCER, Y.C. EL-TAHA, N. KAJUMBA, J.P. MARANGOS, J.W.G. TISCH, M. GABRIELSEN, R.J. COGDELL, E. SPRINGATE, E. TURCU, *Instantaneous Mapping of Coherently Coupled Electronic Transitions and Energy Transfers in a Photosynthetic Complex Using Angle-resolved Coherent Optical Wave-mixing*, in: «Physical Review Let-

ters», vol. CII, 2009, Art. Nr. 057402; E. COLLINI, C.Y. WONG, K.E. WILK, P.M. CURMI, P. BRUMER, G.D. SCHOLES, *Coherently Wired Light-harvesting in Photosynthetic Marine Algae at Ambient Temperature*, in: «Nature», vol. CDLXIII, n. 7281, 2010, pp. 644-647; R. HILDNER, D. BRINKS, J.B. NIEDER, R.J. COGDELL, N.F. VAN HULST, *Quantum Coherent Energy Transfer Over Varying Pathways in Single Light-harvesting Complexes*, in: «Science», vol. CCCXL, n. 6139, 2013, pp. 1448-1451. About warm quantum effects discovered in the avian compass see Z.B. WALTERS, *Quantum Dynamics of the Avian Compass*, in: «Physical Review E», vol. XC, n. 4, 2014, Art. Nr. 042710.

⁴⁰ Most important is David Papineau's famous *Causal Argument for Physicalism* (D. PAPINEAU, *Thinking about Consciousness*, Clarendon Press, Oxford 2002).

⁴¹ R.C. BISHOP, *The Hidden Premiss in the Causal Argument for Physicalism*, in: «Analysis», vol. LXVI, n. 1, 2006, pp. 44-52, here p. 45.

⁴² *Ivi*, p. 47.

⁴³ E.J. Lowe presents a similar thesis in his *Personal Agency* (cit., part I) where he lists the different presentations the Causal Closure Principle has had in the literature and argues that the weaker versions are much more plausible than the stronger ones. He also contends that physicalism is an «unwarranted dogma» based on the faith in the empirically unfounded claim that «no physical event has a non-physical cause» (p. 40). About associated arguments regarding the conservation principles see also S. GIBB, *Closure Principles and the Laws of Conservation of Energy and Momentum*, in: «Dialectica», vol. LXIV, n. 3, 2010, pp. 363-384.