Biol Philos (2009) 24:283–300
DOI 10.1007/s10539-008-9135-9

# Bio-agency and the problem of action

**J. C. Skewes · C. A. Hooker**

Received: 28 June 2008 / Accepted: 4 September 2008 / Published online: 20 September 2008


**Abstract** The Aristotle-Kant tradition requires that autonomous activity must originate within the self and points toward a new type of causation (different from natural efficient causation) associated with teleology. Notoriously, it has so far proven impossible to uncover a workable model of causation satisfying these requirements without an increasingly unsatisfying appeal to extra-physical elements tailor-made for the purpose. In this paper we first provide the essential reason why the standard linear model of efficient causation cannot support the required model of agency: its causal thread model of efficient causation cannot support the core requirement that an action is determined by, and thus an expression of, the agent's nature. We then provide a model that corrects these deficiencies, constructed naturalistically from within contemporary biology, and argue that it provides an appropriate foundation for all the features of genuine agency. Further, we provide general characterisations of freedom and reason suitable to this bio-context (but that also capture the core classical conceptions) and show how this model reconciles them.

**Keywords** Biological autonomy · Bio-agency · Action · Compatibilism

J. C. Skewes
Centre for Functionally Integrative Neuroscience, Aarhus University Hospitals & Department of Philosophy, University of Aarhus, Aarhus, Denmark

C. A. Hooker (✉)
School of Humanities & Social Sciences, Discipline of Philosophy, University of Newcastle, McMullin Bld., Callaghan, NSW 2308, Australia
e-mail: cliff.hooker@newcastle.edu.au

## Introduction

The most powerful traditional frameworks for explaining agency have seemed to require recourse to causation different from the efficient causation of natural science. For instance, Aristotelians differentiate living from inanimate beings by their power to be causes of themselves, while Kantians require that autonomous activity originate from a power for teleological self-determination. Notoriously, it has so far proven impossible to uncover a workable model of causation satisfying these requirements without appealing to extra-physical elements tailor-made for the purpose. The result is that once powerful traditional frameworks for explaining agency have become increasingly unsatisfying as our scientific knowledge of the world has expanded. One response to this has been an increasing demand that agency be naturalised, that is be integrated into natural science.

Fortunately, we shall contend, modern biology, as increasingly underwritten by dynamics of complex systems, holds the key to a new formulation of agent causal power that satisfies the requirements of genuine agency while appealing only to physical (efficient) causation, thus naturalising agency. We introduce that formulation in the next section. But it is essential that we begin by removing inadequate presumptions about causation derived from pre-complex system models. So to prepare the ground we briefly critique the causal-thread model of efficient causation commonly presupposed in non-teleological explanations of agency, in order to identify its inherent incapacity to address the issue satisfactorily.

The causal-thread model essentially pictures causality as a sequence of causally connected events (a thread), with each event in the sequence causing the next. Causal threads can multiply intersect, so that several preceding events may contribute to causing any one event and any one event may contribute to causing several succeeding events. Correlative to this conception is the causal box model of an agent, where an agent is considered a site (the box) into, through, and out of which pass congeries of causal event-threads. In this conception a causal action by an agent is an event-thread of caused changes extending from within the box to a change of state in the external world.

To understand what purchase this model may offer on capturing causation that originates with an agent, consider tracing from an action-caused external state change in the world back in time along the causal thread into the agent causal box. There are two possibilities: either at least one backward branch of the sequence terminates in the box, or all backward branches originating in the box are traceable through it and out into preceding conditions in the world. In the first case the requirement of the action originating with the agent is met but because the initiating event arises ex-nihilo, there is no way for the agent to control the action or for it to be systematically related to the agent's circumstances, so the idea that it is inherently an expression of the agent is undermined. In the second case the action may be an expression of the agent insofar as it is appropriately causally related to agent internal and external conditions but, since all of these are in turn causally determined by preceding events, there is no purchase for the idea of it being an action originating with the agent. We thus need to look beyond the box for a model of agent powers. In doing so we must give up the widespread post-Humean

assumption that causes are events or conditions and treat organisms as genuine holistic loci of causal(-like) power.

This should not surprise since linear causal-thread models break down more generally for all complex dynamical systems that involve simultaneous interactions under multiple feedback. In them causal event-threads have neither beginning nor ending and so offer no resolution of causal responsibility. These problems are especially acute for those systems involving self-organised process architectures and globally organised feedback loops, as living organisms do. Such systems still show causal-like powers in their capacity to do dynamical work in the world (and in themselves) that alters dynamical conditions.[1] In contrast to event-thread analysis, for a century regulatory systems analysts such as control engineers have been successfully dynamically analysing these systems into process architectures with regulatory loci. We shall see that when agents are modelled, as science reveals, in terms of organised complexes of process closures, and with actions correlatively modelled as agent-directed processes, we still obtain a conception of agency with causal powers, but of a radically different non-linear class, that explains why and how agent causality genuinely originates within agents and expresses their natures.

## Towards a process account of agency (I): basic autonomy[2]

The first requirement is to form a conception of agents as sets of processes, in order to distinguish them from their environment and identify them as centres of agency with causal-like power. For this we assume naturalistically that agents are complex (biochemical) dynamical systems. Though they interact with their environment, such systems are distinguished from it by their internal dynamical organisation which is such that a thermodynamically stable entity emerges (see below).

There are two very different ways this can be realised. A quartz crystal is a simple, strongly cohesive object in most environments. Because of its high-energy internal bonding interactions, most perturbing interactions create vibrations within it that are dissipated without disrupting its molecular lattice. Only a sufficiently energetic perturbation, like a hammer blow, will disrupt its molecular lattice,

[1] These complex systems reveal how clumsy is the usual talk of causality, there being so many situations where nothing corresponds to the simple connections it requires. It is better to consider causes as special cases of dynamical processes, viz. those where sufficient energy is transferred to induce a change of state in sufficiently separable system constituents to be identifiable, and to consider dynamics as offering the more general language and criteria. See further, e.g., Hooker (2004).

[2] The following sections grow out of a cluster of linked work on the fundamental characterisation of living organisms based on the bio-organisational notion of autonomy, and includes Bickhard (1993, 2000, 2002, 2005), Bickhard and Terveen (1995), Christensen and Bickhard (2002), Christensen and Hooker (2000a, 2000b, 2002, 2004), Collier (2000, 2004), Hooker (2002, 2008a), Moreno and Ruiz-Mirazo (1999), Moreno and Lasa (2003, Moreno and Etxeberria (2005). This work takes its inspiration from (1970s essays reprinted in) Fong (1996), Varela (1979), Maturana and Varela (1980) and Rosen (1985), but there are important differences, e.g. in the way Maturana and Varela and Rosen emphasise interaction closure while Christensen and Hooker emphasise interactive openness. Late in drafting this paper we became aware of the unpublished work of Campbell (2008a, b) that shares important features with that presented here and with Bickhard, though developed to re-interpret the notion of truth. We commend his work.

destroying its identity. The crystal possesses a passive, static stability: it will persist at equilibrium in the absence of interaction. As such, it is unsuitable as a model for agents. Note that the polar contrast to the crystal, the random gas, is an equally unsuitable model because it has too little internal cohesion to form any interesting internal character at all.

However there is a third alternative: active, because far-from-equilibrium, dynamical stability. This occurs when a system takes energetic input from the environment to maintain its internal cohesive interactions. For instance, a candle flame is a far-from-equilibrium entity which is stable only because of the continual vaporised wax and oxygen inputs that maintain its flame, the resulting temperature and light emission distinguishing its internal thermodynamical state from its environment. For its operation it must remain open to its environment, drawing in resources, utilising them (irreversibly) in its internal processes to regenerate its flame while expelling the waste products, thus contributing to its own stable operation. Thus it is partially self-maintenant; a different, much more active, mode of being.

Living beings are among these open, irreversible, partially self-maintaining systems. They require continuing inputs of oxygen, nutrients and water to regenerate their internal processes, including their capacity for self-maintenance. In this manner they too maintain an internal condition distinct from their environment. But they are self-regenerating to a much higher degree than a candle flame, which has no self-regulatory capacity. For instance, should the flame die down it cannot cause more oxygen and wax vapour to flow in to revive it. Contrast hungry animals actively searching for food to revive themselves. Furthermore, organisms possess a metabolism: by utilising simple primitive chemical inputs they can synthesise all their remaining components and, inserting these at the right locations, repair their bodily processes.

This overall active organisation of processes can be schematised as two cycles, an internal metabolic interaction cycle and an external environmental interaction cycle. These need to be coordinated: the environmental interaction cycle must deliver energy and materials to the organism in a usable form at the times and locations the metabolism requires to complete its regeneration cycle. These two synchronised cycles, jointly producing system regeneration, is the broadest functional characterisation that picks out all and only living individuals, from cells to multi-cellular organisms and, though the detail, especially the dynamical boundaries, vary in graded ways, also many multi-organism formations, such as space stations and cities. In anticipation of its later significance for agency, we shall call this functional biological condition *autonomy*.

Unlike candle flames, autonomous systems actively regulate their own processes. In self-regenerating they are active in synthesis, repair and waste excretion. And most are equally active in their pursuit of self-maintenance, e.g. hunting for food. Even a single cell regenerates itself metabolically from its intake of chemicals through its membrane and can chemotax up, and tumble at random to avoid moving down, a sugar gradient, partially regulating its experience of its environment and its capacity to maintain itself. Multi-cellular animals perform the same overall tasks, but to match their expanded regenerative requirements they do so with an expanded

range of regulatory capacities for both internal and external interaction. The evolution of these latter capacities leads to an increasing capacity to alter their environment to suit resource, safety and breeding needs (e.g. migrate, build mouse holes, farm fungus)—something far beyond the regulatory capacity of the candle.

Living systems, in contrast to candles, are largely self-regulating, requiring a complex, powerful and subtle internal organisation of processes. For instance, the single cell alone is the site of 3000+ chemical processes, each supplied resources by, and supplying products to, other processes. The dynamical organisation of these processes identifies the cell itself as the regulatory locus of the chemotax/tumble, regenerative, and ingestion/expulsion capacities. All living organisms are in this way marked by a strong regulatory asymmetry between themselves and their environment: the locus of living process regulation lies distinctively and substantially within them and not in their environment. Birds organise twigs to make nests, but twigs themselves have no tendency to organise nests or birds.

## Towards a process account of agency (II): self-directedness

This regulatory asymmetry grounds the individuation of living systems in their own dynamical process coordination. The biological conception of autonomy thus captures the root sense of self-governance, the leit motif of the better known but more abstracted socio-political applications of the notion of autonomy. Taken together with the global closure character of regeneration, autonomous systems evidence a distinctive integrated wholeness or global integrity that differentiates them from their environment. This most basic feature is also crucial for properly grounding their agency: it is the basis for insisting that the proper referent for other agency characteristics is the whole agent. It is the whole amoeba that tumbles, the boy that runs. Here we escape the corrosive influence of the causal-thread/causal-box models.

Moreover, autonomous systems direct required resources toward their input gates (e.g. mouths) and internally they direct the flow of energy and materials into the reconstitution of the system (metabolism), both processes shaped to fit the circumstances obtaining. For instance, hunting varies with prey kind and metabolic processes vary with the regeneration required. It thus emerges that an autonomous system (organism) has the fundamentals required for a will, viz. a capacity to do directed work (transmit energy) in relation to the self whose will it is. In sum, an autonomous system is a distinctive individual (its self-regulatory locus) possessed of an active, systematically directive, efficacious capacity for the fundamentals of wilful action.

But this conception can be further enriched. Autonomy is a global constraint; the entire system of processes must so interlock as to regenerate the whole. This multiple process interdependence provides a way of understanding the emergence and functioning of inherent norms. The viability envelope of the system is the range of conditions (e.g. for hydration) under which the process network constitutive of the system succeeds in being self-generating. Since this determines the conditions for continued existence as that system, indicators of these conditions, such as hunger

(glucose deficit) and thirst (hydration deficit), that act as their functional proxies in regulating process modification (including behaviour), are acting in a normative role and indeed constitute the most basic norms available to organisms. More sophisticated systems can also construct new surrogate norms in response to their experiences. For example a young cheetah learns from hunting failure (root norm: hunger) not to break cover too soon, acquiring an operational nearness norm for hunting. And these systems can also modify normative relationships, such as suspending unconditional hunger-driven chasing to creep closer. And they may utilise more general norm signals, such as generalised discomforts and pleasures that provide normative orientation to larger functionally integrative aspects of autonomy.

In combination, this matrix of embodied norm signals—specific, constructed, relationally modified and general—allows a system to direct and evaluate its interaction processes with respect to their autonomy-value for the system. Organisms then co-ordinate their component processes within their viability envelopes to navigate the satisfaction of their norm matrices. This provides them with an elaborated normative perspective for interacting with the world and a corresponding sense of self-integrity.[3] It also equips them with a corresponding root discriminative capacity. Situations are differentiated by their possibilities and norms, the basic judgement taking the form 'It is possible to do activity A in this situation and thereby satisfy norm N'. Thus conceived, activity is inherently anticipative: it anticipates its performance sequence plus the subsequent norm satisfaction as its feedback.[4] Moreover, beyond this level of continuous online *basic anticipation* that lies at the core of all actions, sufficiently sophisticated organisms are also capable of *preparatory anticipation*: they can anticipate the outcomes of certain of their actions 'off-line' (cf. Grush 1997) when preparing for and selecting them. It affords creatures a capacity for the kind of temporally extended goal-directedness that is central to agency.

The activities of autonomous systems are already self-directing, in the basic sense of selectively channelling energy into autonomy satisfaction. However, sufficiently sophisticated organisms can also shape their behaviour, from their normative perspective, to suit their circumstances. The basic dimensions to this are the capacities to (i) dynamically anticipate their interaction processes, (ii) evaluate the outcome of interactions using normative signals and (iii) modify interaction in

[3] Here we have provided a fundamental ground for the emergence of norms in a world of facts. Of course a much larger story has to be told to capture the rich normative life we humans enjoy. For naturalists this will have to be a constructivist and realist story, in something like the way science is. For elements of this story see Bickhard (2002, 2005, 2006), Hooker (1995, Chaps. 5 and 6).

[4] These anticipations provide the foundation for the emergence of truth and representation. An anticipation, say that an interaction is available in the current environment that will satisfy norm N, may succeed or fail when tried; it is, in effect, an implicit predication about that environment that doing A will in fact satisfy N. This constitutes the emergence of a primitive bearer of truth value, the central emergent property of representation. Cf. Bickhard (1993), Campbell (2008a, b). The prediction will hold if the environmental processes presupposed by the possibility and effectiveness of the action—its external dynamical presuppositions (Bickhard 2000)—hold and it is these that provide a powerful implicit content for the anticipation. Bickhard argues, e.g., that this latter provides a solution to the frame problems. See further Bickhard (1993), Bickhard and Terveen (1995).

the light of (i) and (ii) to obtain or improve autonomous closure. Organisms with this three-factor shaping capacity are *self-directed*; directed because they systematically modify their process organisation, and self-directed because the locus of evaluation, the normative perspective driving modification, and the regulatory processes executing it, all lie primarily within the organism itself.

Successful adaptive shaping is problem solving and these three factors are the ingredients from which cognition is formed. Increasing intelligence is expressed in increasingly powerful forms of self-directedness, driven by an increasing capacity to anticipate the longer-term outcomes of actions, to prepare for action through increasingly elaborate preparatory anticipation and, following feedback from action, to elaborate the norm matrix as well as modify behaviour. This elaboration of capacity culminates in self-directed anticipative learning, a capacity to modify self-directedness so as to learn to solve radically new, vaguely specified, open problems—the epitome of human learning and found most developed in science. With such powerful preparatory anticipatory options open to them, these organisms possess a high-order regulatory capacity for fluid goal-directed planning and management of interaction with their environment. That is, they display a powerful intentionality dual to that of intelligence.[5]

In sum, autonomy grounds the central features of agency: possession of a distinctive sense of self grounded in identity-constituting process regulation; an active, systematically directive, efficacious capacity for the fundamentals of wilful action; a normative perspective from which actions are evaluated and that they are anticipated to satisfy; and powerful dual capacities for an intentional grip on situations and to radically learn about them (e.g. learn new behaviours, methods, values) so as to improve self satisfaction. This provides a powerful sense of agency and the basic framework for understanding the relationship between agent, action, and environment. We proceed to develop a general model of agent action in this framework.

## An anticipative process account of action

Typically, maintaining autonomy requires coordinated fluid transitions among internal and external activities. In a complex world organism features can play several different roles (legs can run, jump, kick …) and the organism environment can offer several different opportunities/threats (food, predators, mates …). It becomes the organism's problem to match its capacities to its opportunities and risks moment-to-moment so as to continually satisfy its autonomy. And it may need

[5] Here intentionality is modeled in the spirit of Merleau-Ponty's 'grip' (Merlau-Ponty 1962). The integration of intentionality and intelligence as dual aspects of high-order directed interaction management offers both a more unified conception of mental organisation and a more plausible conception of its evolution than does the traditional split picture in which intentionality is characterised by a language-like referential capacity and intelligence by formal problem-solving capacity. See e.g. Christensen and Hooker (2000b) for some material on the rise of self-directedness and 2002 for the dual integration of intentionality and intelligence, Christensen (2004) for an introduction to integrated executive control, and Farrell and Hooker (2007a, b)on the same processes in science itself.

to shift its plans, on timescales from long to short (e.g. and respectively, seasonal change, storm, predator intrusion), or in response to signals that may vary from wholly anticipatable to complete surprises. In short, generating adaptive action is more like continuously modulating an extended process, rather than assembling complex activity from individually well-formed scripts. A successful hunt, with predator anticipating or responding to prey avoidance tactics, is best conceived as a flexible constraint tube that begins with an open search and slowly narrows to a kill, rather than an algorithm for a sequence of pre-fixed moves. Adaptiveness involves modelling the way the system manages, by modulating its actions, the interaction patterns that are generated. Intelligence is a particular type of modulatory management strategy.

The viability envelope and its related norm matrix constitute the appropriate framework for characterising an integrated agent of this kind. This is because each provides a globally coherent but locally permissive constraint: any activity is permissible so long as the constraint is satisfied. A viability envelope specifies a general set of constraints and allows continuously varying combinations of bodily activities to be selected within that envelope as a function of the continuously changing context. Mammals, for instance, are not confined to fixed hunger-satiation and predator fear values but can continuously vary their risk level and derived precautionary food-searching values with urgency of need, size of apparent reward, nearness of threat and so on; that is, as a function of context. Similarly, a norm matrix specifies a general set of evaluation dimensions and allows continuously varying combinations of values, including constructed values, to be selected within the matrix as operative according to the changing context. This allows for the kind of fluid context-sensitive variation we see in creatures and that real environments demand.

Thus to characterise an action tube the agent needs to differentiate or delineate two complementary aspects: the relevant performance constraints and the relevant norms. The relevant performance constraints are those aspects of the viability envelope that pertain to the operative context. For a hunt they would include the hunter's capacity to select prey, avoid injury, and so on. There will also be a corresponding web of conditional constraints among these, e.g. that you can stalk closer upwind, but not select prey well if the sun is in your eyes. The relevant norms will include the end goal(s), specified in terms of the dominant norm(s) being pursued overall (e.g. nutrient satiation), along with all the derived norms that will apply to the hunting process because of the preceding performance constraints (e.g. intact, untorn foot pads). There will also be a corresponding web of conditional priorities among these determining such key aspects as when and how to modify or abandon a hunt because the prey is too skilful at avoidance or injury threatens too greatly.

This complex characterisation must be anticipated about the hunting action tube if the organism is to be properly prepared. The relation of the agent's dominant to derived norms (e.g. satiation to initiating nearness for hunting) is therefore one of increasing recursiveness of the anticipations involved. For derived norms, each action is carried out first anticipating that these norms will be satisfied as their environmental objects are encountered (e.g. the cheetah creeps, trying to get nearer)

and second anticipating that meeting derived norms (nearness) will satisfy the dominant norm (satiation) while failure to meet them will contribute to frustration of the dominant norm. Similar considerations apply to preparedness to abandon the hunt mid-way, e.g. because of injury threat. A rough representation of the end goal of the action tube must therefore be established at the outset of the action as a preparatory anticipation of its outcome, to play the role of the dominant norm, with the structure of the action tube characterised by anticipation of satisfying conditions for its derived norms, the whole modulated relative to this by the micro online basic anticipations associated with neuro-muscular coordination of activity. The agent thereby sets up long sighted preparatory anticipations that modulate lower online basic anticipations associated with ongoing near-term behaviours, the whole being something that can properly be called a goal directed action.

This evaluative-anticipative potentiating process is the initial internal process that engages agents with their situation in relation to their own processes, the necessary precondition for their anticipatively shaping future outcomes through their activity. An immature (self-directed) organism is marked out by having inadequate anticipations of these and other of their niche-defining activities. While it has basic micro online behavioural anticipations (it can run, bite, etc., though perhaps clumsily), it cannot yet form the complex preparatory anticipations necessary for stable end goals.

Setting up for a particular kind of action (e.g. a hunt) involves both specifying the anticipated relevant performance features and norms and the relevant webs of interdependencies among them. This is especially complex because each component process typically assumes others are successfully proceeding, e.g. that visual perception formation will inform muscle innervation in a way that will support leaping at fleeing prey. Let us call that particularised part of the viability envelope that is engaged for the occasion the contextual performance envelope, the particularised part of the norm matrix the contextual performance norm matrix, and call the interrelated collection of presupposed performances underlying these the contextual web of internal dynamical presuppositions. All three of these are part of anticipative preparation for a context-sensitive performance, at whatever competence level bodily organisation demands and permits. Call this set-up process *anticipative potentiation*, making a particular anticipation active or potent.

We shall take anticipative potentiation to include the generation of all the interrelated potentialities among the various modes of activity that facilitate smooth 'in-flight' transitions among them and conversely the potentialities for transitioning from the particular mode (say a hunt) to other organisational modes (say cub protection), according as the situation offers/threatens. These potentialities equally need preparing if the organism is to transition smoothly to another mode of activity while executing hunting mode—again, at whatever competence level bodily organisation demands and permits.[6]

[6] Bickhard, following some neurophysiological usage, uses the term microgenesis to describe the part of this process that occurs in the central nervous system and is confined to short-term potentiating of the current activity (e.g. the hunt). However, autonomic and hormonal potentiation are more ancient than CNS potentiation and remain fundamentally important to action even in species where intelligence is of great functional importance, hence we have chosen a more encompassing term.

We are now in a position to characterise action within this framework. Anticipative potentiation sets up the constituting framework for action, for there is an anticipative comprehension of the performance situation that gives intentional focus its grip, and a complementary evaluation of performance from a normative perspective that gives intentional content its grip, so that we indeed have a fully fleshed out agency at work. An action, then, is any activity carried out as part of an orchestrated autonomous organism response to a situation that is framed by anticipative potentiation. The sense of action (along with intention and evaluation) is weakest when only basic anticipation is involved and grows stronger and richer as preparatory anticipation (along with derived norms and learning capacity) grows stronger and richer.

## Reconciling freedom and rationality of action

Our basic work is complete, we have developed a naturalised notion of agency and action in which an agent's actions have efficient causal power, and agent causality genuinely originates within agents and expresses their natures. To show the merit of this approach we now briefly indicate how it provides a natural resolution to the venerable and vexed problem of how actions can be both free and reasonable.

Before commencing we pause to remind ourselves of why the causal box conception of agents is incapable of this resolution. From this perspective it is only possible to model freedom as not being wholly determined by preceding external causes and reasonableness as a species of internal causal coordination of internal and external circumstances. The conflict between them is already incipient in this characterisation. In the case where the relevant causal sequence terminates in the box the action may have a claim on being free on the grounds that it is under-determined by preceding external causes, but there is no way for the agent to control the action or for it to be systematically related to the agent's circumstances and so no purchase for the idea of acting reasonably. In the case where the relevant causal sequence is traceable through the agent-box and out into preceding conditions in the world the action may have a claim on being rational, on the grounds that the agent-box is the site of the conjoining of causal threads that relate to the agent's internal and external circumstances, but since the action sequence is wholly causally determined by preceding events there is no purchase for the idea of the agent acting freely.

There is a crudity of formulation forced by assuming the causal-box model of agents. Free action, defined as externally under-determined action, defines freedom in terms of simple events rather than agent characteristics, neglecting the internal functioning of agents that makes them free. Similarly, while reasonable action, defined as internal causal coordination of internal and external circumstances, is more apt, it too fails to consider the nature of the internal coordination. While the notions of reasonableness and freedom are too complex to explore in any detail here, we begin our alternative account by presenting certain general characterisations of them that are more focused on the nature of agency itself.

### Reasonableness

We presume that the purpose of a concept of reasonableness is to limn the boundaries of agency coordination of norm, belief and action. Actions that are at least normatively relevant to an agent's life concerns and are shaped in the light of relevant agent beliefs thereby possess an inherent rationale for the agent and provide a sense of cohering unity to agency. Systems possessing too little such unity, for whatever reason, cease to be candidates for agency.

### Freedom

We presume that the purpose of a concept of freedom is to limn the boundaries of agency responsibility. Agents can freely act so as to make a difference in this world and are held responsible for their actions when, and because, their influence derives peculiarly from their own agency and expresses its character.

This suffices for our purposes here. We add that these attributions apply to all organisms, according to their capacities, and that from a biological perspective reasonableness is rooted in a fundamental bio-organisational capacity, one that under-pins both cognitive and intentional capacities,[7] making it the more fundamental attribute. An efficient and, (more powerfully) an efficacious, agent seeks to improve internal coordination, thus exhibiting stronger forms of cohering unity. Reason denotes the internal organisation of coordination processes in agents capable of at least improving efficiency.[8]

It is then natural to model freedom as a regulatory condition on agents' interaction processes with their environment, and reasonableness as a regulatory condition on agents' internal process organisation, including feedback from interaction with their environment. And then, we shall contend, a clearer picture emerges of freedom and reason as conditions of agents and any incipient principled clash between freedom and reason is avoided.

What grounds an agent's choices is the permissive nature of its viability envelope and norm matrix that typically permit their joint satisfaction in many different ways. The integrity of agent permissive options is guaranteed by the agent's dynamical organisation as a locus of self-regulation. While it is often easy to remove some choices, it is hard to remove them all. (An isolated prisoner can still sing to herself.) As external signals create their perturbations, the agent's internal regulatory response is to preserve its autonomy, whence changing environmental or internal conditions typically opens up some choices even if it also closes off others. (Cases of environmental forcing, as when hit by a tsunami or injected with anaesthetic, do not present counterexamples, simply disruptions of agency regulatory capacity to varying degrees.) And this discretionary capacity increases with increasingly

[7] For discussion see Christensen and Hooker (2002).

[8] Agent efficacy also concerns promoting the development of agency itself and hence allows for the re-conception of norms, beliefs, alternatives and efficiency criteria (see e.g. Hooker (1995, Chap. 6); cf. Brown (1988). Nonetheless, all finite bio-agents also have severe limits on their capacity to be efficient (see Cherniak 1986; Hooker 1994). Nor is there any founding commitment to logic in natural reason, let alone in reasonableness (see e.g. Hooker 1994, 1995, Chaps. 5, 6; cf. Brown 1988).

sophisticated capacity for preparatory anticipation. Thus, the regulatory constitution of agents ensures that under normal circumstances they will always have the permissive space, defined by their viability envelope and norm matrix, to resolve their preparatory state into a specific anticipative potentiation, so that in constructing their own preparatory anticipations they can self-regulate the nature of this potentiation. Each such potentiation choice occurs inside the self-regulatory locus constituting the agent and in that proper sense distinctively originates with the agent.

Such choices also manifestly coordinate norms, beliefs, and action. The focusing of permissive range down on to a specific anticipative potentiation is normally brought about by the intersection of the internal norm signals currently active and the agent's ongoing receipt of sensory signals indicating the kind of situation it is currently in. Such agent commitments are as richly shaped by the agent's internal models of how such situations potentially unfold—their potentia maps—as agent sophistication permits.[9] And as resolutions of the agent's full norm matrix, actions richly express the agent's normative perspective, ensuring that they are normatively relevant to the agent's life concerns. Thus agent actions possess an inherent rationale and provide a sense of cohering unity for the agent. They are thus inherently reasonable and, assuming they are defeasibly efficient (note 8), they are fully rational.[10]

Thus the autonomy conception of agency grounds conceptions of reason and freedom as mutually interlocked and developing capacities, both enabled by determinism, with interactive norms central to both, a fundamental aspect of the wider integration of intentionality and intelligence as two complementary integrative capacities (note 5). This provides a powerful perspective on the evolution of self-regulation and a correlative developmental perspective on the emergence of individual selfhood.

Nor while agent regulatory integrity is in place can the causal(-like) effects of external signals be followed through an agent in the event-thread manner because the multiple interconnecting process loops constituting the agent renders that dynamically inoperable. After a short time each of the many components (however identified, but especially if structurally identified) have contributed to re-constituting so many other components and processes through feedback and feedforward that causal thread attribution breaks down. Rather, a system/environment regulatory asymmetry is established expressing a global metabolic self-regenerative closure. Autonomous agents are in that sense dynamically emergent entities for whom reduction fails.

Nor should we let the general requirement of determinism daunt us here: determinism says only that in fact each complete dynamical state is followed by a unique successor. This is compatible with the dynamical reality of regulatory

[9] We so easily speak of beliefs here, but it is worth emphasising with Bickhard that the fundamental form they take in all organisms is that of feature-interaction potentia maps, with elements of the kind 'Sensory features F indicate that action A is possible and, if C, leads to feature G and satisfaction of norm N, and …'. It is a constructive work of great sophistication to 'compile' such schemas into the subject-object belief syntax we humans take for granted (cf. Christensen and Hooker 2000b, Part IV).

[10] See Hooker (1995), especially Chaps. 5, 6; cf. Bickhard (2002).

asymmetry (as simple deterministic controller-system behaviour demonstrates) and hence compatible with genuine dynamical possibility of choice for the agent. That is, given that agent identity is constituted in its process and corresponding functional organisation, determinism is consistent with it following from agency that a variety of different regulatory responses be dynamically possible and the actual one chosen on any occasion is resolved substantially or wholly within the agent's regulatory locus.[11]

## Reconstructing philosophical traditions

While the reconciliation of freedom with reason sufficiently shows the latent power of the approach, we conclude by briefly indicating its capacity to capture central aspects of other philosophical traditions. First, we briefly elaborate the relationship to central elements of Merlau-Ponty's philosophy, for which we have already briefly noted a sympathetic relationship at note 5. We then briefly show how the approach even captures central features (but not the metaphysics) of classic libertarianism, here represented by Chisholm (1976).

The primary aim of Merleau-Ponty's philosophy is to understand behaviour as a genuine product of organism agency. Merleau-Ponty (1963) starts out by rejecting as inappropriate for this purpose the two models of behaviour assumed by then contemporary neuroscientists in explaining the role of the nervous system in organism motility. These are the reflex arc model, according to which behaviour is a physiological elaboration of sensory stimuli in the action domain, and the central sector model, according to which behaviour is produced by neocortical processes co-ordinating sensory inputs to construct co-ordinated action processes. Merleau-Ponty's argument is that in the first case behaviour must be a product, however indirectly, of the state of the external environment and not simply of an internal reflex, while in the latter behaviour must be an expression of genuinely integrated agency and not simply a summed product of discrete higher level neocortical processes. Our approach shares Merleau-Ponty's starting point, since Merleau-Ponty's arguments against the capacity of these models of behaviour to capture genuine agency amount to a specific version in the domain of neuroscience of our general rejection on similar holistic grounds of the causal box model and advocacy of integrated autonomy bio-agency.

[11] In one standard terminology (McKenna, Michael, "Compatibilism", The Stanford Encyclopedia of Philosophy (Summer 2004 Edition), Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/sum2004/entries/compatibilism/), in a deterministic world self-regulation is a special form of guidance that exhibits the internal dynamical interrelationships requisite for constituting dynamical regulation. The mere requirement of guidance fails to discriminate agency from dynamical activity generally—everything makes some difference to dynamical state sequence—while the autonomous form of dynamical organisation provides for a central role to anticipation characteristic of agency. Being unencumbered to choose an alternative is not sufficient for good regulatory control. Being able to match present circumstances with different conditional outcomes on the basis of internal processes such as motivation is also required, which requires some capacity for forward modelling and the formation of context-dependent action plans. Anticipative processes estimate at $t_2$ the consequences of choosing different alternatives at $t_1$, providing important information for constraining regulative control at $t_1$.

Merleau-Ponty's positive response is to treat behaviour as an emergent dynamical whole that has neither agent internal processes nor the external environment as its sole cause. Rather, the agent is understood as having a non-decomposable internal functional organisation modulated by "lines of force" imposed by the environment. In addition, the agent is understood as playing an active role in this modulation, since the lines of force it encounters are in turn largely determined by its behavioural history and its own internal organisation. As a framework, this is a less specific version of our suggestion that the agent be treated as an integrated complex system differentiated by its internal dynamics, which are partly shaped in response to an external world to which it stands in an asymmetrical regulatory relation.

Within this framework Merleau-Ponty (1962) develops an understanding of action as an extended dynamical process which could be initiated and or modulated by the agent, and where the initiation and modulation of action is understood as being driven by the agent's implicit anticipations of how the dynamical details of the action unfold. This is a more general description of the model of action in terms of anticipative potentiation we have developed above. The general understanding of agency we have developed here is thus in important ways consonant with the philosophical understanding offered by Merleau-Ponty.

It might be objected that our account, committed to a complex systems ontology and hence a kind of naturalism, is incompatible with Merleau-Ponty's phenomenological foundations, in particular with the sincere reservations about the viability of naturalism which he took to follow from those foundations. However, Merleau-Ponty's reservations had as their target Humean naturalism, which diverges considerably from a naturalism based on the complex systems ontology we assume (see "Introduction"), and Merleau-Ponty's objections do not apply unmodified to the latter. At the very least, and as Thompson observes (Thompson 2007), Merleau-Ponty's commitment to understanding agency as a dynamical process is entirely compatible with complex systems analysis as a mathematical and explanatory engine. Thus, even if there emerged fundamental disagreement between our own approach and an approach that is more faithful to Merleau-Ponty's phenomenological foundations, both could learn much from one another.

With respect to Chisholm, who felt libertarianism unavoidable if the central features of agency are to be captured, it is clear, first, that an autonomous agent manifests an inherently clear sense of individuality, as Chisholm assumes. An agent is individuated by its self-regulation. Indeed, it has as rich an individuality as the combination of its inherited distinctiveness and its subsequent unique context-dependent learning bestow upon it, and this much flows inherently from its nature as an autonomous system. As does, second, its possession of an inherent normative perspective from which it acts and evaluates the consequences of action. Thus far the general perspective; now to action.

Third, autonomous agents are inherently active and their activeness has causal (-like) power, for each can do dynamical work in the world to bring about the delivery of requisite resources and do work internally to regenerate itself. Thus are satisfied the first two of Chisholm's requirements for agent activity: that agents can

and do engage in causal(-like) changing of the world and themselves (p. 199).[12] Finally, the regulatory constitution of agents ensures that in constructing their own preparatory anticipations agents can self-regulate the nature of this potentiation. This ensures satisfaction of Chisholm's third requirement that there be the genuine possibility of an alternative action—that while an agent may do A it would also be true that "at the time there was something else he [the agent] could have done instead" (p. 199). Thus do Chisholm's metaphysics lose their libertarian point since the claims about agency and action they were intended to underpin can be captured within the autonomy dynamical conception.

Chisholm elaborates a language for understanding action causally in terms of 'direct' action—the initial internal change within an agent that is the undertaking of an action. In the autonomy model anticipative potentiation is the clear candidate for direct action since it is the initial internal change in an agent that precedes the behavioural expression of any action, yet constitutes the most basic part of undertaking one because it frames the execution of any directed activity that follows, including the observable behavior that identifies the action for others.

And this gives the same properties to undertaking as Chisholm demands, most notably that (i) (p. 208) undertakings contribute causally to bringing about the anticipated external change (i.e. potentiation regulates bodily activities that ultimately do work in the external world); (ii) (p. 206) if an agent causes an external change then it undertakes some prior condition (i.e. agents must first potentiate and drive bodily change from it); (iii) (p. 206) an agent contributes causally to its own undertakings (i.e. agents are the dynamical origin of their potentiations); (iv) (p. 209) there are 'basic' actions, successful direct undertakings or "things we succeed in doing without undertaking still other things to get them done", e.g. bodily activity, for most people, in normal circumstances (i.e. potentiation and the subsequent processes within the agent fall within the locus of the agent's self-regulation and (ii) holds); and (v) (p. 208) undertaking is intentional, whence an agent can undertake an activity yet fail to succeed in changing the world as intended (i.e. potentiation is intentional qua anticipative, and can fail to culminate in the intended external goal at any subsequent stage through failure of dynamical presuppositions).

## Conclusion

The autonomous bio-agent model, developed within a regulatory dynamical metaphysics and properly elaborated, is both scientifically based and useful, and philosophically constructive and powerful. Because autonomous dynamical unity and regulatory asymmetry are real dynamical properties, autonomy supports a natural order of norms and action that delivers strong forms of self-regulation compatibly with determinism—indeed requires determinism for efficacy. It is then natural to model reasonableness as a regulatory condition on an agent's internal process organisation and freedom as a regulatory condition on an agent's interaction

[12] Hereafter all unidentified page references are to Chisholm (1976).

processes with their environment. This grounds conceptions of reason and freedom as mutually interlocked and developing capacities, both enabled by determinism, with interactive norms central to both, forming aspects of the wider bio-organisational integration expressed in the dual capacities of intentionality and intelligence. The possibility of irreversible far-from-equilibrium global self-regulatory bio-organisation makes a significant metaphysical difference to the world, for it creates the objective possibility of integral agency.

The shift from simple linear thread models of causality to contemporary complex systems dynamical models, poses several distinctive challenges to philosophical understanding. One locus for these is the set of challenges thrown up by the new systems and synthetic biology, the complex systems successors to molecular genetics. These concern understanding biological organisation in general and global organisational constraints like autonomy in particular, for which we presently have no established general mathematical treatment, and the ways in which they are realised through interrelated congeries of biochemical processes (the biosynthetic pathways). Kaufman (2000) suggests the link goes through what he calls work-constraint cycles, where dynamical work done in one process creates the constraint framework for other processes (cf. water flow cutting river banks), and something like this seems right. But precisely how such ideas are to be integrated, and whether they provide any basis for a principled decomposition of total functional capacity into sub-capacities and/or of global process into modules of the sort that is central to both current scientific methodology and traditional philosophical analysis, remains unclear.[13] The other locus of philosophical challenges concerns the need to re-draft our conception of agency in complex systems terms. This both concerns our understanding of the basic causal powers of agency and its general capacities—like freedom and rationality—and its limitations—from akrasia to finitude—and also its specific biological embodiment. One important aspect of this latter is re-thinking cognitive neuro-science from within the complex systems framework, where the elegant but simplistic divide between functional and process analyses—cf. software and hardware—that sustained computational AI and its philosophical framework is being radically re-thought. This domain poses many of the same challenges as are currently posed to cellular biology. The present essay and its cited supporting papers constitute some first steps in responding to re-thinking agency.

**Acknowledgement** Mark Bickhard and Richard Campbell are thanked for insightful and supportive comments on earlier drafts.

## References

Bickhard M (1993) Representational content in humans and machines. J Exp Theor Artif Intell 5:285–333. doi:10.1080/09528139308953775

Bickhard M (2000) Autonomy, function, and representation. In: Etxeberria A, Moreno A, Umerez J (eds) Contribution of artificial life and the sciences of complexity to the understanding of autonomous systems, communication & cognition, vol 17, special edition, pp 111–131

[13] For initial exploration of many of the issues, see Fu and Hooker (2008), Hooker (2008b) and references.

Bickhard M (2002) Critical principles: On the negative side of rationality. N Ideas Psychol 20:1–34. doi: 10.1016/S0732-118X(01)00010-1
Bickhard M (2005) Interactivism: a manifesto. http://www.lehigh.edu/~mhb0/
Bickhard M (2006) The social ontology of persons. http://www.lehigh.edu/~mhb0/pubspage.html
Bickhard M, Terveen L (1995) Foundational issues in artificial intelligence and cognitive science—impasse and solution. Elsevier Scientific, Amsterdam
Brown H (1988) Rationality. Routledge, London
Campbell R (2008a) A process-based model for an interactive ontology. In: Bickhard M (ed) *Synthese*, special issue on interactivism (to appear)
Campbell R (2008b). Doing truth (in preparation)
Cherniak C (1986) Minimal rationality. MIT Press, Cambridge Mass
Chisholm R (1976) The agent as cause. In: Brand M, Walton D (eds) Action theory. D. Reidel, Boston, pp 199–211
Christensen W (2004) Self-directedness, integration and higher cognition. Lang Sci 266:661–692. special issue on distributed cognition and integrationist linguistics. doi:10.1016/j.langsci.2004.09.010
Christensen W, Bickhard M (2002) The process dynamics of normative function. Monist 851:3–28
Christensen W, Hooker C (2000a) Organised interactive construction: the nature of autonomy and the emergence of intelligence. In: Etxeberria A, Moreno A, Umerez J (eds) Contribution of artificial life and the sciences of complexity to the understanding of autonomous systems, communication & cognition, vol 17, special edition, pp 133–158
Christensen W, Hooker C (2000b) An interactivist-constructivist approach to intelligence: self-directed anticipative learning. Philos Psychol 13(1):5–45. doi:10.1080/09515080050002717
Christensen W, Hooker C (2002) Self-directed agents. In: MacIntosh J (ed) Can J Philos, 31, 19–52, special supplementary volume on contemporary naturalist theories of evolution and intentionality
Christensen W, Hooker C (2004) Representation and the meaning of life. In: Clapin H, Staines P, Slezak P (eds) Representation in mind: new approaches to mental representation. Elsevier, Sydney, pp 41–69
Collier J (2000) Autonomy and process closure as the basis for functionality. In: Chandler J, van de Vijver G (eds) Closure: emergent organisations and their dynamics. Ann N Y Acad Sci 901:280–291
Collier J (2004) Interactively open autonomy unifies two approaches to function. In: Dubois D (ed) Computing anticipatory systems: CASY03 sixth international conference, AIP conference proceedings, vol 718. American Institute of Physics, Melville, pp 228–235
Farrell R, Hooker C (2007a) Applying self-directed anticipative learning to science. I: agency and the interactive exploration of possibility space in Ape language research. Perspect Sci 15:86–123
Farrell R, Hooker C (2007b) Applying self-directed anticipative learning to science. II: learning how to learn across 'revolutions'. Perspect Sci 15:220–253
Fong P (1996) The unification of science and humanity. New Forums Press, Stillwater
Fu P, Hooker C (2008) Outstanding issues in systems and synthetic biology. In: Fu P, Latterich M, Panke S (eds) Systems biology and synthetic biology. Wiley, New York (to appear)
Grush R (1997) The architecture of representation. Philos Psychol 10(1):5–25. doi:10.1080/09515089708573201
Hooker C (1994) Idealisation, naturalism, and rationality: some lessons from minimal rationality. Synthese 99:181–231
Hooker C (1995) Reason, regulation and realism: toward a naturalistic, regulatory systems theory of reason. State University of New York Press, Albany
Hooker C (2002) An integrating scaffold: toward an autonomy-theoretic modelling of cultural change. In: Wheeler M, Ziman J (eds) The evolution of cultural entities. British Academy of Science and Oxford, Oxford UP, pp 67–86
Hooker C (2004) Asymptotics, reduction and emergence. Br J Philos Sci 55:435–479. doi: 10.1093/bjps/55.3.435
Hooker C (2008a) Interaction and bio-cognitive order. In: Bickhard M (ed) Synthese, special issue on interactivism (to appear). doi: 10.1007/s11229-008-9374-y
Hooker C (2008b) On fundamental implications of systems and synthetic biology. In: Fu P, Latterich M, Panke S (eds) Systems biology and synthetic biology. Wiley, New York (to appear)
Maturana H, Varela F (1980) Autopoiesis and cognition. Reidel, Dordrecht
Merleau-Ponty M (1962) Phenomenology of perception (trans: Smith C). Routledge & Kegan-Paul, London
Merleau-Ponty M (1963) The structure of behavior (trans: Fisher A). Beacon Press, Boston

Moreno A, Etxeberria A (2005) Agency in natural and artificial systems. Artif Life 11(1–2):161–175. doi: 10.1162/1064546053278919
Moreno A, Lasa A (2003) From basic adaptivity to early mind: the origin and evolution of cognitive capacities. Evol Cogn 19:12–30
Moreno A, Ruiz-Mirazo K (1999) Metabolism and the problem of its universalisation. Biosystems 49: 45–61. doi:10.1016/S0303-2647(98)00034-3
Rosen R (1985) Anticipatory systems: philosophical, mathematical, and methodological foundations. Pergamon, New York
Thompson E (2007) Mind in life: biology, phenomenology, and the sciences of mind. Harvard University Press, Cambridge
Varela F (1979) Principles of biological autonomy. Elsevier, New York