Metaphysical Conservatism and Mechanical Characteristics of Human Nature

Brian J. Dellinger and Jon Doyle
Department of Computer Science
North Carolina State University
Raleigh, NC 27695-8206
BJDelli2@ncsu.edu, Jon_Doyle@ncsu.edu

Abstract

We sketch some reasons for thinking that modern formulations of mechanics might shed some light on old questions about human nature, in particular the malleability of human nature and the character of limits on cognition and volition. We illustrate the approach by showing how to regard certain forces that depend on mental states or the states of artificial controllers as interaction forces between minds or controllers and their environments, thus allowing physical measurement of some mental properties.

Introduction

People have long disagreed about whether human nature is fixed or flexible, and to what degree and in what ways. The great monotheisms, for example, regard humans as essentially flawed no matter what, while some ideologies have regarded humans as perfectible with proper nurture or environment. These disagreements about perfectability do not extend to disputing the ability of people to learn and unlearn things, or to make use of natural or designed artifacts to change human abilities. Artificial intelligence plays a mixed role in answering these questions, agreeing with limits on unaided human abilities (commonly rationality and computability), but offering a possibility for overcoming these limits by giving humans new tools. The question then becomes whether human nature changes as people adopt new technologies.

It is clear that new technologies can change the things people do in ways that strengthen or weaken natural and learned abilities. Postman (1985; 1992), for example, argued that properties of the information communication medium offered by television skew content toward passive entertainment and distraction, and predicted the displacement of reading and degredation in ability for sequential thought. Gelernter (2013) observes related changes wrought by the internet on writing and the content of writing. From an economic point of view, changing technology changes the utility landscape faced by the human, leading to changing choices in how humans spend their time and changing sets of skills learned and unlearned in the new distribution of activities (compare (Nelson and Winter 1982)). One might still regard these changing abilities as leaving invariant a central core of fixed tastes and rational calculation (Stigler and Becker 1977), but the information needed to characterize how properties of the environment satisfy these tastes plays a role in *homo economicus* analogous to the role the input machine description plays in a universal Turing machine. Indeed, Turing's (1936) mathematical model of human calculating behavior and its distillation in the notion of a universal computer shows that the question of fixity and flexibility cannot be answered on the basis of overt fixed structure alone, as universal machines can be so tiny as to seem essentially trivial, for example Minsky's (1962) four-symbol seven-state universal machine and Wolfram's two-state three-symbol machine. In this setting, fixity concerns the form and content of the specified behavior rather than of the means by which these specifications are followed.

We seek to understand to what extent one can characterize and measure the ability of people to change by using concepts from mechanics to augment those of computation and economics. Can one measure the force and work needed to produce a given deformation in one's beliefs or desires? Can one identify mental elasticity, plasticity, and rigidity in persons, and measure the force required to maintain a desired deformation? Can one measure the forces of habit and distraction generated by deformations that require counteracting force and effort? To what extent do bounds on available forces explain differences in effort between learning and unlearing?

To pursue this line of investigation requires treatment of people as objects in realm of mechanics. This has been a contentious issue since the time of Descartes, who proposed that persons are composed of both a mental substance forming the mind and a physical substance forming the body. Even the nascent natural philosophy of the time proved embarrassing for this proposal, and by the time calculus and laws of mechanics were formalized mathematically, there seemed to be no way to fit Cartesian mental substances into what appeared to be a fruitful theory of physical substances. Today, mechanics is commonly regarded as a theory of physical substances alone. That view is mistaken.

The development of the modern axioms of continuum mechanics by Noll (1973) and Truesdell (1991) in the last century undermined this common conception of mechanics as strictly limited to the physical realm. Developing axioms that apply to the wide variety of physical substances produces a formalization in which the physical nature of mechanical objects appears only in embedded assumptions that space and time form continua, and in which the existence of specific physical forces appears only in laws of special material types. Reformulating the axioms to remove assumptions of a continuum nature of space and time, as was accomplished in (Doyle 2006), preserves the central character of mechanics, and allows one to form mechanical systems as hybrid combinations of mechanical subsystems. This allows, for example, forming mechanical systems that include both physical bodies existing in continuum space and time and mental bodies existing in somewhat discrete space and time, and in which the laws of special physical materials are complemented by laws of special types of mental materials.

One virtue of this extended formalization of mechanics is that it permits one to study human nature from a position of metaphysical conservatism or openness in which one avoids dualism, monism, physicalism and other metaphysical assumptions at the start and distinguishes properties of human nature that can be characterized in terms of mechanical properties alone from properties that require adoption of additional metaphysical assumptions.

The development of mechanical characterizations of human nature, mental flexibility, and the difficulty of mental change is in its infancy (Doyle 2006; 2010; 2013), and we cannot present a nonspeculative characterization at this time. We instead illustrate the approach by sketching how one might use a metaphysically conservative conception of mechanics to translate the unexplained conception of control or agency found in some extant notions of hybrid systems (Branicky 2005; Branicky, Borkar, and Mitter 1998) and hybrid automata (Alur et al. 1993) into equivalent mechanical interactions between controllers and physical bodies. Work in progress builds on the approach of metaphysical conservatism to analyze mental inertia, effort, and the material properties mentioned previously.

Agency and Interaction

Traditional cognitive approaches to psychology understand the nature of interactions between persons and their environments in terms of natural vocabularies of human agency—belief, desire, intent, deliberation, volition, and actions at many levels of detail and significance—but say little about how these cognitive elements are realized in the body or connect to the world in which the agent acts, and so say little about how embodiment might shape mind (Minsky 1965; Newell 1982). Other approaches use physical and neurophysiological vocabularies— neurons, connections, energy states, and dynamical systems at different levels of granularity—but provide little guidance on how to connect physiological dynamics to ordinary intentional explanations of behavior, and few principles for designing rather than learning or evolving human-level cognition. In between these approaches lie hybrid automata formalizations (Alur et al. 1993) in which a discrete automaton interacts with a continuum dynamical system by stipulated rules of unspecified physical or computational character.

We believe that mechanical concepts permit one to connect such views of mind and body without requiring lengthy reductions of one set of terms to the other. The following shows how to translate some physical effects of the actions of agents into mechanical interactions between each agent and the world. This translation allows one to measure and compare forces exerted by different controllers on the same body, irrespective of assumptions about the physicality of the controllers.

We limit attention here to physical effects describable as changes in the forces that physical bodies exert on each other, which we call *agency* forces, and to mechanical interactions consisting of forces between agent and world, which we call *interaction* forces. We introduce a notion of *mediation* that yields explicit formulas for translating agency forces into interaction forces. This translation preserves the resultant forces acting on each body and thus the ordinary mechanical behavior. The translation retains the identity of the agent as a producer of change but does not require assumptions about the physicality of the agent. We further limit attention to showing how to identify forces exerted by agents in the case that physical forces depend on agent states. The structure or nature of agent states will not be at issue in the present analysis, and we will use the relatively neutral term "controller" rather than "agent," "person," "mind," and "body." We will use the term "body" as a general term of mechanics, and do not intend to convey some of the common connotations of the other terms about what they include or exclude. For simplicity of exposition, we treat only instantaneous forces, which suffice to demonstrate our main construction and which stand independent of assumptions about the origins, locality, or determinism of forces.

Plan of the Paper

We first summarize the modern formalization of a system of forces over a universe of mechanical bodies and extend the ontology of bodies to include controllers, as well as systems of forces over these extended universes.

We next formalize the notion of agency seen in variable-state controllers and state-dependent systems of physical forces among physical bodies and contrast agency forces with interaction forces between controllers and physical bodies. We identify notions of controller independence and additivity that capture natural structural and behavioral properties of agency in a multi-controller setting.

Finally, we show how to regard controllers as mediating changes in forces between physical bodies. We apply mediation to construct a state-dependent system of forces over the enlarged domain of controller and physical bodies, one that yields the same resultant forces on the original physical bodies as in the original conditional force system. We also identify conditions under which the constructed interaction forces agree with the original agency forces in purely physical universes, and briefly discuss the issues involved in reconstructing the original physical-domain force system from the hybrid-domain transformation.

Space limitations preclude presentation of most proofs and subordinate results. None of the proofs are particularly difficult, but some involve lengthy algebraic manipulations. The subordinate results provide stronger but more complicated versions of the results presented here.

Mechanical Ontology

Mechanics extends geometry with nongeometric concepts including those of body, force, mass, and energy. Mechanics is based on general and special laws defining these concepts and governing their interrelations. For example, mass is not a completely general mechanical concept, but appears as the generator of the inertial and gravitational force in special laws. The present treatment follows the approach of (Doyle 2006) but involves only the notions of bodies, forces, and possibly nonmechanical conceptions of controller states.

We begin with a set Ω of bodies that forms a Boolean lattice with respect to the *subbody* or *part-of* relation $\sqsubseteq$, with meet $\sqcap$ and join $\sqcup$, and *null body* $\bot$ and *universal body* $\top$. We write $\overline{B}$ to denote the *environment* or *exterior* of a body B, and say that bodies B and C are *separate* iff $B \sqcap C = \bot$.

It is common to regard bodies as sets of *body points*, indeed, as subsets of the set of points constituting the universal body $\top$, so that $C \sqsubseteq B$ implies $C \subseteq B$ for each $B, C \in \Omega$. Continuum mechanics considers bodies made up of continua rather than isolated body points, but none of the treatment that follows requires making any assumption about whether bodies form sets or continua; only the Boolean lattice structure will matter.

An instantaneous *system of forces* or *force system* on a universe Ω is an assignment

$$f : (\Omega \times \Omega) \to \mathcal{V} \tag{1}$$

to all pairs of bodies of Ω of force values. Mechanics typically considers force values as elements of real vector spaces, but we here assume only that $\mathcal{V}$ forms a commutative additive group with null element 0. We read $f(B, C)$ as the *force exerted on* B *by* C in the force system f. We call $f(B, \overline{B})$ the *resultant force on* B, that is, the force exerted on B by its exterior. Mechanics requires that force systems be *null-passive*, meaning that only a null or zero force is exerted on or by the null body $\bot$, and *additive* in each argument on separate bodies, or more generally, for all C_1 and C_2,

$$f(C_1 \sqcup C_2, B) = f(C_1, B) + f(C_2, B) - f(C_1 \sqcap C_2, B)$$
$$f(B, C_1 \sqcup C_2) = f(B, C_1) + f(B, C_2) - f(B, C_1 \sqcap C_2).$$

Finally, the mechanical force systems considered here also must be *pairwise equilibrated* on separate bodies, meaning that all mutual forces are equal and opposite, or formally, that

$$f(B, C) = -f(C, B)$$

for all separate bodies B and C.

We write $\mathcal{F}(\Omega, \mathcal{V})$ to denote the set of all systems of mechanical forces over Ω and $\mathcal{V}$. This set of force systems forms a commutative additive group by defining addition and additive inverse of force systems pointwise. In the present treatment, we will consider only the case in which $\mathcal{V}$ consists of ordinary physical force values, written $\mathcal{V}_{\mathrm{p}}$.

Controller Ontology

Within the universe Ω, we distinguish a *physical* subuniverse Ω_{p} with universal body $\top_{\mathrm{p}}$, and a *controller* subuniverse Ω_{c} with universal body $\top_{\mathrm{c}}$. We thus regard controllers as mechanical bodies rather than as nonmechanical objects. We make no assumptions at this time about the extent or overlap of Ω_{p} and Ω_{c}. That is, writing Ω_{pc} to denote the subuniverse common to Ω_{p} and Ω_{c}, we allow that Ω_{pc} might be empty apart from $\bot$ (as Descartes might have it) or might be all of Ω (as Spinoza might have it).

If $B \in \Omega_{\mathrm{p}}$, we write $\overline{B}^{\mathrm{p}}$ to denote the *physical exterior* of B within Ω_{p}, and if $C \in \Omega_{\mathrm{c}}$, we write $\overline{C}^{\mathrm{c}}$ to denote the exterior of C within Ω_{c}, called the *controller exterior* of C.

Restricting each system of forces in $\mathcal{F}(\Omega, \mathcal{V}_{\mathrm{p}})$ to physical bodies yields the set $\mathcal{F}(\Omega_{\mathrm{p}}, \mathcal{V}_{\mathrm{p}})$ of force systems over physical bodies.

In addition to bodies and forces, the conception of agency studied here assumes that the forces acting in the physical system depend on *states* inhabited by controllers. We write $\Psi(C)$ to denote the (nonempty) set of possible states of C, and assume that each *universal* or *global* state $\psi \in \Psi(\top_{\mathrm{c}})$ of the universal controller body both determines and is determined by the states of all other controller bodies, writing $C(\psi)$ to denote the state that C inhabits within ψ.

Agency and Interaction Forces

We characterize controller-dependent forces by defining a *conditional force system over* Ω_{p} *and* $\mathcal{V}_{\mathrm{p}}$ to be a mapping of the form

$$f[\cdot] : \Psi(\top_{\mathrm{c}}) \to \mathcal{F}(\Omega_{\mathrm{p}}, \mathcal{V}_{\mathrm{p}}). \tag{2}$$

We write $f[\psi](A, B)$ to mean the *conditional force* of B on A in state ψ of C.

We call conditional force systems of the form (2) *agency* forces because in such a system the controller plays the role of the agent of change in physical forces between bodies. This conception of agency, however, requires nothing about how the controller plays this role or about the nature of the controller. If we wish to regard controller agency as mechanical agency, it is natural to expand the conception of controllers to hybrid-domain force systems of the form

$$h[\cdot] : \Psi(\top_{\mathrm{c}}) \to \mathcal{F}(\Omega, \mathcal{V}). \tag{3}$$

We call these *interaction* forces because they involve forces between controller bodies and physical bodies.

Of course, some controllers might be physical, but the minimal conception of agency forces captured in (2) makes no assumptions about the physicality of controllers or the existence of forces between controllers as such and physical bodies. We may thus regard controllers for which no physical realization is specified, such as the automata portion of hybrid automata, as having a nonphysical or informational character that allows separate attribution of controller interaction forces to controllers as such rather than confounding them with interaction forces that are naturally attributed to physical realizations of controllers but that play no significant role in control activities. For example, the force of gravity acting on an aircraft flight computer adds load to the wings, but one does not consider this element of the force exerted by the controller as an aspect of its control activities. The notion of interaction forces permits such separation between forces of control and forces of physical embodiments.

The following shows how to recast forces over $\Omega_{\rm p}$ as interaction forces over Ω for systems in which controllers exhibit some natural forms of independence. For such systems, we construct a system h of type (3) that preserves the resultant forces on physical bodies holding in the system f, so that for each $A \in \Omega_{\rm p}$ and $\psi \in \Psi(\top_{\rm c})$ we have

$$h[\psi](A, \overline{A}) = f[\psi](A, \overline{A}^{\rm p}) \qquad (4)$$

or equivalently, for all $A \in \Omega$, as

$$h[\psi](A \sqcap \top_{\rm p}, \overline{A \sqcap \top_{\rm p}}) = f[\psi](A \sqcap \top_{\rm p}, \overline{A} \sqcap \top_{\rm p}). \qquad (5)$$

In the present treatment, we take preservation of resultant forces to constitute mechanical equivalence of force systems. This condition reflects the central principle of mechanics that forces of all origins contribute in the same way to the overall force exerted on a body. Such uniformity of treatment applies as well to the self-forces bodies exert on themselves, but mechanics typically assumes all self-forces vanish, in which case the total force on a body reduces to just the resultant force. As we will see, the construction presented here in fact preserves self forces as long as they do not vary with controller states, a condition that applies in the case of vanishing self forces. We have constructions that transform systems with nonzero self forces into mechanically equivalent systems with no nonzero self forces, but do not these present here.

Controller Independence

One element common to conceptions of human and automatic agency is the idea that such agents exert influences largely independent of others. Some notion of independence underlies engineering applications, as in aircraft flight controllers that control only one aircraft and do not manipulate the parts of other aircraft. It also underlies familiar notions of freedom of action or will of persons, as when one person makes decisions and takes actions independently of the decisions and actions of others. We do not assume that the same conceptions of independence are plausible for both engineered controllers and human minds. To simplify presentation of the main results, however, the following assumes that separate controllers can consist and act completely independently of each other. Our results also obtain under weaker assumptions.

We assume first that each controller C is *state independent* in the sense that its state can vary independently of the states inhabited by separate controllers. Formally, for each $\psi \in \Psi(C)$ and each $\overline{\psi} \in \Psi(\overline{C}^{\rm c})$ we assume that there exists in $\Psi(\top_{\rm c})$ a state we write as $(\psi, \overline{\psi})$, such that $\psi = C((\psi, \overline{\psi}))$ and $\overline{\psi} = \overline{C}^{\rm c}((\psi, \overline{\psi}))$. We abbreviate $f[(\psi, \overline{\psi})]$ as $f[\psi, \overline{\psi}]$.

We say that a controller C is *force independent (of its exterior) with respect to* f to mean that changing the state of C from ψ to ψ' produces the same changes in forces regardless of the unchanged state of C's complement, or formally,

$$f[\psi, \overline{\psi}] - f[\psi', \overline{\psi}] = f[\psi, \overline{\psi}'] - f[\psi', \overline{\psi}']. \qquad (6)$$

for each $\psi, \psi' \in \Psi(C)$ and $\overline{\psi}, \overline{\psi}' \in \Psi(\overline{C}^{\rm c})$.

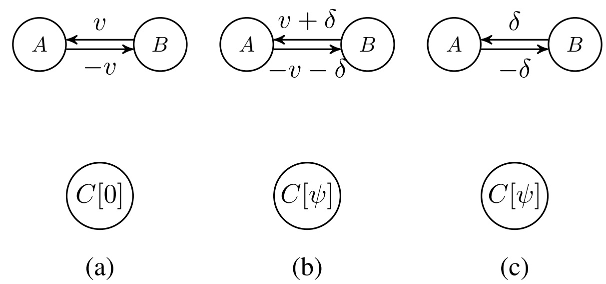


Figure 1: Variation in forces between A and B as the state of controller C varies. In (a), the global state is the null state 0, and $f[0](A, B) = v$. In (b), the global state ψ is one that agrees with 0 on $\overline{C}^{\rm c}$ and in which $f[\psi](A, B) = v + \delta$. (c) depicts the same state as (b), but here the force values denote the differential force $f^*[C, \psi](A, B) = f[\psi](A, B) - f[0](A, B)$.

We say that a conditional force system f displays *force independence* just in case every controller in $\Omega_{\rm c}$ is force independent with respect to f.

Controller Additivity

To identify the effects or contributions made by controllers to the overall system of forces, we consider the differences in forces that result as a controller changes state, holding constant the states of separate controllers. If these controllers are force independent, these differences in forces will not depend on which states the other controllers inhabit. Accordingly, we may regard some state in $\Psi(C)$ as being the *null* or *base* state of controller $C \in \Omega_{\rm c}$. We assume there exists a global null state 0 in which each controller C is in its null state $C(0)$, and will overload the symbol 0 by abbreviating $C(0)$ as 0. Nothing in the following depends on the means by which one identifies null states. For engineered controllers, null states typically will consist of the "off" state; for minds, null states might be taken to be ones in which the individual is not acting volitionally.

Let f be a conditional force system of the type (2) and 0 a null state chosen from $\Psi(\top_{\rm c})$. The overall change of forces $f[\psi] - f[0]$ due to a change of state from 0 to ψ constitutes the difference attributable to all controllers. Our focus now is on identifying the difference in forces attributable to change of state of each individual controller. We define the *differential force system based on f and* 0 to be the mapping f^* of the form

$$f^*[\cdot, \cdot] : \Omega_{\rm c} \times \Psi(\top_{\rm c}) \to \mathcal{F}(\Omega_{\rm p}, \mathcal{V}_{\rm p}) \qquad (7)$$

defined by

$$\begin{aligned} f^*[C, \psi] &\stackrel{\rm def}{=} f[C(\psi), 0] - f[0] \qquad (8)\\ &= f[C(\psi), \overline{C}^{\rm c}(0)] - f[C(0), \overline{C}^{\rm c}(0)]. \end{aligned}$$

We illustrate this notion in Figure 1. In particular, we have

$$f[\psi] = f[0] + f^*[\top_{\rm c}, \psi].$$

Because $f^*[C, \psi]$ is the difference between two force systems, it also is a force system.

One can compare the differential forces of different controllers to obtain a range of measures of strength. For example, let C_1 and C_2 be controllers and B a physical body, so that $f^*[C_1, \psi](B, \overline{B})$ and $f^*[C_2, \psi](B, \overline{B})$ denote corresponding changes in the resultant force on B. In the instantaneous setting considered here, one might compare these effects by their magnitudes or directions, and might say that C_1 is stronger than C_2 if the magnitudes of change due to C_1 are at least as great as those due to C_2 for all physical bodies. In a temporal setting, one can compare rates of work done by the different controllers as well.

Identification of differential force systems allows one to restate the notion of force independence in terms of additivity of changes. We call an individual controller C *controller-additive with respect to* f^* just in case

$$f^*[\top_{\mathrm{c}}, \psi] = f^*[C, \psi] + f^*[\overline{C}^{\mathrm{c}}, \psi] \tag{9}$$

holds in every state $\psi \in \Psi(\top_{\mathrm{c}})$. We say that f^* is controller additive if every controller is controller additive with respect to it. We say that f is controller additive if f^* is controller additive for every choice of 0. Our independence assumptions make controller additivity equivalent to force independence.

Theorem 1. *If all controllers are state independent, a conditional force system is force independent if and only if it is controller additive.*

Mediating Interaction Forces

We now show how to construct a hybrid-domain conditional force system of the form (3) from a physical-domain conditional force system of the form (2) that preserves resultant forces on physical bodies in the sense of (5).

The simplest hybrid-domain extension of a physical-domain system f just matches the values of f on physical bodies and assigns zero forces to purely nonphysical bodies, which can be expressed by the definition

$$\bar{h}[\psi](A, B) \stackrel{\mathrm{def}}{=} f[\psi](A \sqcap \top_{\mathrm{p}}, B \sqcap \top_{\mathrm{p}}). \tag{10}$$

With this definition, the restriction of $\bar{h}$ to Ω_{p} is just the original system f. The system $\bar{h}$ clearly preserves resultant forces on physical bodies but just paraphrases the lack of interaction between nonphysical controllers and physical bodies already present in the original force system f. This adds no insight, and one might as well rest content with f.

To obtain nontrivial mechanical interactions between controllers and physical bodies, we make use of differential force systems defined over the full universe of bodies, with

$$h^*[\cdot, \cdot] : \Omega_{\mathrm{c}} \times \Psi(\top_{\mathrm{c}}) \to \mathcal{F}(\Omega, \mathcal{V}) \tag{11}$$

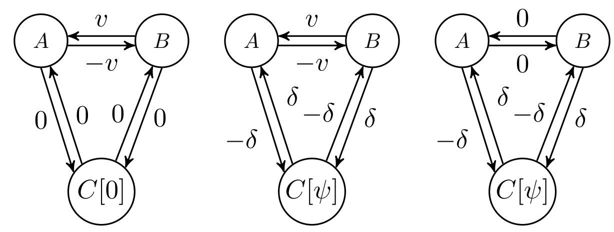
satisfying

$$h^*[C, \psi] = h[C(\psi), 0] - h[0] \tag{12}$$

and

$$h[\psi] = h[0] + h^*[\top_{\mathrm{c}}, \psi]. \tag{13}$$

Figure 2, which depicts physical bodies A and B and separate nonphysical controller C, illustrates the intuition underlying our construction. We regard changes in forces from



(a) (b) (c)

Figure 2: Variation in hybrid forces between A and B as the state of a nonphysical controller C varies, corresponding to the three cases displayed in Figure 1. In (a), controllers are in the null state 0, and $h[0](A, B) = v$. In (b), controllers are in a state ψ in which $h[\psi](A, B)$ remains v but C adds the force δ to the force on A via $h[\psi](A, C)$ and subtracts the force δ from the force on B via $h[\psi](B, C)$. (c) depicts the same state as (b), but here the force values denote the differential hybrid forces $h^*[C, \psi]$.

those existing in the null state as forces exerted by the controller on the physical bodies, with forces exerted on one physical body the opposite of the force exerted on the other physical body. We call these forces *mediating interactions* because each controller mediates changes in forces between physical bodies, something as if the controller were a rigid body interposed between the physical bodies so as to transmit contact forces unchanged. Formally, we say that a differential force system h^* of type (11) *mediates external force changes* for a differential force system f^*, or more briefly, h^* *mediates* f^*, just in case h^* satisfies the following three conditions (14), (15), and (16).

The first two conditions require that changes to forces between bodies due to a controller must go through the controller. Suppose $A, B \in \Omega$ and $C \in \Omega_{\mathrm{c}}$, with A and B each separate from C. The first condition requires that h^* should reassign changes in resultant forces on A due to changes in the state of C to C itself, or formally, that

$$h^*[C, \psi](A, C) = f^*[C, \psi](A \sqcap \top_{\mathrm{p}}, \overline{A} \sqcap \top_{\mathrm{p}}). \tag{14}$$

Because C is in the exterior of A, this reassignment at most shifts the sources of forces from one body in this exterior to another body in the exterior, leaving the resultant force acting on A unchanged. The second condition requires that h^* exert zero differential forces between A and B, or formally, that

$$h^*[C, \psi](A, B) = 0. \tag{15}$$

Put another way, h^* only remaps forces involving controllers, and makes no changes in base forces between bodies separate from controllers.

The third condition limits controllers to changing forces on other bodies, not on themselves. That is, any force exerted by a controller on itself does not change as the states of the controller change, or formally, that

$$h^*[C, \psi](C, C) = 0. \tag{16}$$

Invariance of all self forces under state changes implies this condition, but we do not make that more general assumption here. One might regard the third condition as confirming that a controller cannot mediate changes in its own self force, as this is already a force between the controller and itself.

We say that h mediates interaction forces if h^* does so for every choice of null state. Our main result is the following.

Theorem 2. *Suppose all controllers are state independent, f is a controller additive conditional force system, and f^* is the differential force system based on f and $0 \in \Psi(\top_c)$. Then h is a conditional force system mediating interaction forces for f if and only if*

$$\begin{aligned} h[\psi](A,B) = f[0](A \sqcap \top_p, B \sqcap \top_p) \\ + f^*[B \sqcap \top_c, \psi](A \sqcap \top_p, \overline{A} \sqcap \top_p) \\ + f^*[A \sqcap \top_c, \psi](\overline{B} \sqcap \top_p, B \sqcap \top_p). \end{aligned} \quad (17)$$

Proof sketch. The "if" portion of the proof involves only straightforward verification of the needed properties, and implies that h^* is controller additive because f^* is. To prove the "only if" portion, use $A, B \in \Omega$ to divide $C \in \Omega_c$ into four parts and use controller additivity of h^* to obtain

$$\begin{aligned} h^*[C,\psi] = h^*[A \sqcap B \sqcap C, \psi] + h^*[A \sqcap \overline{B} \sqcap C, \psi] \\ + h^*[\overline{A} \sqcap B \sqcap C, \psi] + h^*[\overline{A} \sqcap \overline{B} \sqcap C, \psi]. \end{aligned}$$

Apply each of the functions in this equation to A and B. The last term vanishes by (15), yielding

$$\begin{aligned} h^*[C,\psi](A,B) = h^*[A \sqcap B \sqcap C, \psi](A,B) \\ + h^*[A \sqcap \overline{B} \sqcap C, \psi](A,B) \\ + h^*[\overline{A} \sqcap B \sqcap C, \psi](A,B). \end{aligned} \quad (18)$$

By (15), the final term of (18) reduces to

$$\begin{aligned} h^*[\overline{A} \sqcap B \sqcap C, \psi](A,B) = \\ h^*[\overline{A} \sqcap B \sqcap C, \psi](A, (\overline{A} \sqcap B \sqcap C)), \end{aligned} \quad (19)$$

the second to last term of (18) reduces to

$$\begin{aligned} h^*[A \sqcap \overline{B} \sqcap C, \psi](A,B) = \\ h^*[A \sqcap \overline{B} \sqcap C, \psi]((A \sqcap \overline{B} \sqcap C), B), \end{aligned} \quad (20)$$

and the first term of (18) reduces to

$$\begin{aligned} h^*[A \sqcap B \sqcap C, \psi](A,B) = \\ h^*[A \sqcap B \sqcap C, \psi](A, (A \sqcap B \sqcap C)) \\ + h^*[A \sqcap B \sqcap C, \psi]((A \sqcap B \sqcap C), B) \\ - h^*[A \sqcap B \sqcap C, \psi]((A \sqcap B \sqcap C), (A \sqcap B \sqcap C)). \end{aligned} \quad (21)$$

But by (16), the third term of (21) vanishes, so combining the remainder of (21) with (20) and (19) yields

$$\begin{aligned} h^*[C,\psi](A,B) = h^*[A \sqcap B \sqcap C, \psi](A, (A \sqcap B \sqcap C)) \\ + h^*[A \sqcap B \sqcap C, \psi]((A \sqcap B \sqcap C), B) \\ + h^*[A \sqcap \overline{B} \sqcap C, \psi]((A \sqcap \overline{B} \sqcap C), B) \\ + h^*[\overline{A} \sqcap B \sqcap C, \psi](A, (\overline{A} \sqcap B \sqcap C)). \end{aligned}$$

Two of these terms are in a form to which we may apply (14) directly, and pairwise equilibration puts the other two into the same form to yield

$$\begin{aligned} h^*[C,\psi](A,B) = f^*[A \sqcap B \sqcap C, \psi](A \sqcap \top_p, \overline{A} \sqcap \top_p) \\ - f^*[A \sqcap B \sqcap C, \psi](B \sqcap \top_p, \overline{B} \sqcap \top_p) \\ - f^*[A \sqcap \overline{B} \sqcap C, \psi](B \sqcap \top_p, \overline{B} \sqcap \top_p) \\ + f^*[\overline{A} \sqcap B \sqcap C, \psi](A \sqcap \top_p, \overline{A} \sqcap \top_p). \end{aligned}$$

Using controller additivity of f^*, we combine the first and fourth terms and the second and third terms to obtain

$$\begin{aligned} h^*[C,\psi](A,B) = f^*[B \sqcap C, \psi](A \sqcap \top_p, \overline{A} \sqcap \top_p) \\ - f^*[A \sqcap C, \psi](B \sqcap \top_p, \overline{B} \sqcap \top_p). \end{aligned}$$

Applying pairwise equilibration to the negated term, substituting $\top_c$ for C, and combining with (13) then yields (17). □

Clearly, $h[0]$ assigns the same forces to bodies as $f[0]$ assigns to their physical parts, that is,

$$h[0](A,B) = f[0](A \sqcap \top_p, B \sqcap \top_p).$$

Moreover, mediating interaction forces preserve resultant forces.

Theorem 3. *If h is a force-mediating extension of f, then h preserves resultant forces in the sense that for each $C \in \Omega_c$ and $A \in \Omega$ we have*

$$h[\psi](A \sqcap \top_p, \overline{A \sqcap \top_p}) = f[\psi](A \sqcap \top_p, \overline{A} \sqcap \top_p) \quad (22)$$

In the case in which all controllers are physical bodies, Theorem 2 suggests a condition under which f and h agree. For simplicity, assume that $\Omega = \Omega_p$, so that f and h have the same domain. In this case, we say f is *self mediating* just in case it mediates external force changes for itself. We then have the following result.

Theorem 4. *If all bodies are physical, all controllers are state independent, and f is a self-mediating conditional force system, then $f = h$.*

Indeed, if all bodies are physical, self-mediation means that $f = h = \bar{h}$. One might regard this result as indicating the inherent conservatism of the interaction-force construction. Even when all bodies are physical, however, self-mediation is clearly a special case, for forces between bodies separate from a controller can vary in a conditional force system but not in a self-mediating one.

Another question of interest is whether we can recover f from h. Each h determines $f[0] = h[0]$ for physical bodies, which leaves recovery of $f^*[\top_c, \psi]$. This is sometimes possible in the case of a purely physical universe, for as noted earlier, if f is self-mediating, then h equals f. Apart from this case, however, we do not have a unique reconstruction, but instead have the following result.

Theorem 5. *If f_1 and f_2 are conditional force systems of the same type that assign the same resultant forces to each body, then $h_1 = h_2$ if and only if $f_1[0] = f_2[0]$.*

Summary and Discussion

On the assumption that separate controllers exhibit certain natural independence conditions, formula (17) expresses our main result, namely a constructive method for converting a possibly mechanically inhomogeneous conception of controller agency into a mechanically homogeneous form of interaction between agents and the world they inhabit. The construction reassigns changes in forces between bodies to controllers involved in the changes in conformity with a natural notion of mediation. It preserves resultant forces on all bodies and does not introduce self-forces where none existed before, and so preserves the basis of mechanical behavior of the controlled system.

The interaction forces we construct vary parametrically with the frame of reference as represented by controller base states. We expect that to engineer intelligent agents, some would seek to obtain the explanatory and intelligibility benefits of cognitive or computational states by constructing or choosing base states that separate cognitive information from noninformational physical states, that align with natural distinctions between forces attributable to the base physical system and forces attributable to cognitive or information-based actions. Nothing in the formalism presented here enforces such a separation, as the mediation notion used to construct h reattributes all force changes to controllers, whether these changes in forces represent physical or nonphysical information state changes. Specifically, under our independence assumptions we may rewrite (17) as

$$h[\psi] = h[0] + h^*[\top_c \sqcap \top_p, \psi] + h^*[\top_c \sqcap \overline{\top_p}, \psi].$$

This expression for h explicitly divides changes in forces due to changes in states of physical controllers, represented by the term $f^*[\top_c \sqcap \top_p, \psi]$, from changes in forces due to changes in states of nonphysical controllers, represented by the term $h^*[\top_c \sqcap \overline{\top_p}, \psi]$. If one regards purely physical parts of controllers as inhabiting only controller state 0 and nonphysical parts of controllers as inhabiting nonphysical informational or mental states, the physical-controller portion of the force vanishes and we have

$$h[\psi] = h[0] + h^*[\top_c \sqcap \overline{\top_p}, \psi].$$

In this case, one can ascribe all changes in forces to state changes of nonphysical controllers.

As noted in the introduction, the results presented here represent only promissory notes on a full mechanical analysis of the properties of minds and a transfer of results of modern continuum mechanics to the mental realm. Preliminary discussion of measuring mental effort in mechanical terms can be found in (Doyle 2013).

Economists tend to regard preferences and expectations as universal means for combining or summarizing all influences on mental behavior. Mechanics plays a similar universal role for physical behavior, but now offers the possibility of treating minds, meanings, and values as mechanical entities, thereby increasing the visible uniformity of the world.

Acknowledgements

Jon Doyle is supported in part by the SAS Institute Distinguished Professorship in Computer Science.

References

Alur, R.; Courcoubetis, C.; Henzinger, T. A.; and Ho, P.-H. 1993. Hybrid automata: an algorithmic approach to the specification and verification of hybrid systems. In Grossman, R.; Nerode, A.; Ravn, A.; and Rischel, H., eds., *Hybrid Systems*, volume 736 of *Lecture Notes in Computer Science*, 209–229. Berlin: Springer-Verlag.

Branicky, M. S.; Borkar, V. S.; and Mitter, S. K. 1998. A unified framework for optimal control: Model and optimal control theory. *IEEE Transactions on Automatic Control* 43(1):31–45.

Branicky, M. 2005. Introduction to hybrid systems. In Hristu-Varsakelis, D., and Levine, W., eds., *Handbook of Networked and Embedded Control Systems*. Boston: Birkhäuser. 91–116.

Doyle, J. 2006. *Extending Mechanics to Minds: The Mechanical Foundations of Psychology and Economics*. London, UK: Cambridge University Press.

Doyle, J. 2010. Toward a quantitative theory of belief change: Structure, difficulty, and likelihood (a progress report). Technical Report TR-2010-20, Department of Computer Science, North Carolina State University, Raleigh, NC.

Doyle, J. 2013. Mechanics and mental change. In Küppers, B.-O.; Hahn, U.; and Artmann, S., eds., *Evolution of Semantic Systems*. Berlin: Springer-Verlag. 127–150.

Gelernter, D. 2013. Worry about internet drivel. In *2013: What *Should* We Be Worried About?* Edge.org.

Minsky, M. L. 1962. Size and structure of universal turing machines using tag systems. In *Recursive Function Theory, Symposium in Pure Mathematics 5*. Providence, RI: American Mathematical Society.

Minsky, M. L. 1965. Matter, mind, and models. In *Information Processing: Proceedings of the IFIP Congress*, 45–49. Amsterdam: North-Holland.

Nelson, R. R., and Winter, S. G. 1982. *An Evolutionary Theory of Economic Change*. Cambridge: Harvard University Press.

Newell, A. 1982. The knowledge level. *Artificial Intelligence* 18(1):87–127.

Noll, W. 1973. Lectures on the foundations of continuum mechanics and thermodynamics. *Archive for Rational Mechanics and Analysis* 52:62–92.

Postman, N. 1985. *Amusing Ourselves to Death: Public Discourse in the Age of Show Business*. Penguin.

Postman, N. 1992. *Technopoly: The Surrender of Culture to Technology*. New York, NY: Knopf.

Stigler, G. J., and Becker, G. S. 1977. De gustibus non est disputandum. *American Economic Review* 67:76–90.

Truesdell, III, C. A. 1991. *A First Course in Rational Continuum Mechanics*, volume 1: General Concepts. New York: Academic Press, second edition.

Turing, A. M. 1936. On computable numbers with an application to the entscheidungsproblem. *Proceedings of the London Mathematical Society Series 2* 42:230–265.