International Journal of Machine Consciousness
Vol. 3, No. 1 (2011) 219−235

DOI: 10.1142/S1793843011000698

HEGELIAN PHENOMENOLOGY AND ROBOTICS

DONALD S. BORRETT[*,‡], DAVID SHIH[†,§], MICHAEL TOMKO[†,¶],
SARAH BORRETT[†,||] and HON C. KWAN[†,**]

[*]*Divison of Neurology, Toronto East General Hospital,*
840 Coxwell Ave. #221, Toronto, Ontario, M4C 5T2, Canada

[†]*Department of Physiology, University of Toronto,*
Toronto, Ontario, M5S 1A8, Canada
[‡]*d.borrett@utoronto.ca*
[§]*david.shih@utoronto.ca*
[¶]*miketomko@hotmail.com*
[||]*8sb7@queensu.ca*
[**]*h.kwan@utoronto.ca*

A formalism is developed that treats a robot as a subject that can interpret its own experience rather than an object that is interpreted within our experience. A regulative definition of a meaningful experience in robots is proposed in which the present sensible experience is considered meaningful to the agent, as the subject of the experience, if it can be related to the agent's temporal horizons. This definition is validated by demonstrating that such an experience in evolutionary autonomous agents is embodied, contextual and normative, as is required for the maintenance of phenomenological accuracy. With this formalism it is shown how a dialectic similar to that described in Hegelian phenomenology can emerge in the robotic experience and why the presence of such a dialectic can serve as a constraint in the further development of cognitive agents.

Keywords: Synthetic phenomenology; evolutionary autonomous agents; Hegelian dialectic.

1. Introduction

Unless a particular behavior is meaningful to a robot and not just the programmer, the robot will never function in a truly autonomous fashion nor be able to make decisions through mechanisms that even remotely resemble the mechanisms by which we make decisions [Di Paolo, 2002; Manzotti and Tagliasco, 2005]. We propose a regulative definition of a meaningful experience in a robot in which the present sensible state is considered meaningful to the agent, as the subject of the experience, if it can be related to the agent's own extended time horizons. This definition is validated by showing that such an experience in evolutionary autonomous agents (EAAs) is embodied, contextual and normative as is required for maintenance of phenomenological accuracy. In addition, it will be demonstrated how a dialectic, similar to that proposed in Hegel's phenomenology, can arise in the agent's experience and why the presence of a dialectic can be used as a constraint in the development of agents whose experience becomes meaningfully organized by the agent itself and not through an external framework imposed by the programmer.

The appeal to Hegelian phenomenology addresses another concern relevant to the design of intelligent robots, the development of autonomous agents that do not rely on the programmer or any outside influence to modify behavior. Although an object is external to a robot, it is assumed that this relationship of externality has to be maintained in dealing with the truth condition of the robot's knowledge of the object during the development of an autonomous agent. This results in the constant need of an external arbitrator to decide if the knowledge of the object that is internalized in the robot corresponds to the object in the external world. Hegel's solution of internalizing both the object and knowledge of the object, what is to be tested and the criteria for knowledge, in consciousness can be extended to robotic design. By internalizing both the object of experience and knowledge of the object in the dynamics of a robotic controller and evolving an agent in which correspondence between these two moments is a constraint on optimal functioning, a truly autonomous agent can emerge that can arbitrate decisions concerning its behavior without any appeal to outside influences.

1.1. *Meaning in robots*

Attempts to instantiate more sophisticated mechanisms in robotic design that mirror our own cognitive abilities usually will take structures that are the outcome of our own interpretive processes such as the concepts of agency, motivation or attention and synthesize an agent that incorporates these structures into its architecture. These structures are usually reproduced by specifying a module in the robotic architecture that functions in a manner that we interpret would be how such a module should function to accomplish the desired task [Sun, 2007]. An example of such an approach is Baars and Franklin's LIDA model of the global workspace theory of consciousness [Baars and Franklin, 2009]. A robotic architecture that uses modules to simulate perceptual associative memory, episodic memory and declarative memory, procedural memory and action selection is established based on how we would interpret the workings of a global workspace theory of consciousness in the robot. With this approach however, the agents incorporate concepts that are the outcomes of our interpretive processes and not necessarily the mechanism by which we interpret our experiences to develop these concepts. By extrapolating the results of our interpretive framework onto an artificial agent, we preclude the possibility of that agent developing its own interpretive framework by which its own contents can become meaningful.

If we cannot use the conceptual contents of our experience to serve as the framework to develop robotic architecture, three directions can be taken to assist in the implementation of the mechanisms by which these contents emerge and meaning is grounded. Firstly, empirical studies that utilize only neuroscientific data can point to mechanisms by which we construct a synthetic reality from sensory motor data [Chemero and Silberstein, 2007]. But neuroscientific data is limitless and unless this data can be organized more concisely, direct extrapolations to robots will be difficult. Secondly, evolutionary algorithms in which survival is the only constraint on population selection avoids any task specific architecture and allows the robots to evolve an architecture independent of any outside control. Although promising, such an approach requires that all the details of function emerge spontaneously, a constraint that compromises any realistic time expectancy for the evolution of such an agent [Floreano and Mondada, 1996]. Thirdly and most relevant to the topic of this paper, phenomenology can provide insights into how our own interpretative mechanisms are structured by reflecting on the form rather than the content of experience. Since the form of phenomenal experience is invariant across all types of experience, mechanism must preserve form if it is to provide a causal explanation for the emergence of content. Put another way, our own phenomenology is a direct reflection of the underlying mechanisms by which our nervous system operates divorced as possible from any formal framework that attempts such an explanation. A robot whose design is based on the underlying mechanisms by which our own experience is organized must reproduce this phenomenology in its own experience. Controversy will obviously arise concerning the particular form that is taken as the most fundamental in our experience. For the purposes of this paper, we take the form of our experience as having the fundamental structure of a present sensible experience that is framed by a temporal horizon. This temporal horizon is either implicit in the experience or explicitly accessed depending on the needs of the subject at that time. It is with this form of phenomenal experience that meaningful experience in a robot, as subject of this experience, is identified and the dialectic is sought.

"Presence in absence" is a concept that characterizes our phenomenal experience and has been considered the most fundamental principle of all phenomenologies [Russon, 2003]. Although our immediate sensible experience is in the present, this experience is meaningful only because of what is absent, the entire temporal context within which it occurs. Within the temporal horizons that frame the present sensible experience are the subject's past history and future expectations and these horizons are "carried with the subject" [Gadamer, 2000] in her sensory motor interactions with the environment. In robotic design, the present always assumes a privileged position regardless of whether a top-down or bottom-up strategy is chosen. However, if a robot is provided details of its entire temporal history including its past and its future expectations, and if it is allowed to use this temporal information to guide its normative evolution, then the privileged position of the present can be removed. When we access details about our past history or future expectations, this may be accomplished by determinate memories or imagined consequences which is woven into the normative framework of our sensory-motor interactions. A robot, rather than having specific memories or expectations, has time scales embedded within the dynamics of its controller. Allowing the robot access to these time scales during its interaction with the environment would therefore enable the robot to instantiate this "presence in absence".

The temporal horizons that frame sensible experience have been equated with the dynamical system theory concept of temporal hierarchical organization (THO) which is a reflection of the nature and distribution of time scales in the signal itself [Borrett *et al.*, 2006]. Instantiation of "presence-in-absence" in the agent was suggested by the feedback of the THO into the robot controller. Possible implementation strategies for this architecture include the simple decomposition of the time signal of the robot controller into its frequency components and the feedback of this frequency information into the controller. Meaningful experience in a robot corresponds to the time evolution of the present sensible experience that is continuously framed by the extended temporal history of the agent which represents the standard or criterion by which the present is interpreted. In dynamical terms, the sensible flux in experience is the time evolution of the states of the controller whereas the extended temporal history is the THO that feedsback into the network itself.

With implementation of this form of experience, the robot possesses the regulative requirement for a meaningful experience. With the establishment of a dialectic within this experience, it will also possess the ability to interpret its own experiences. We use the term "interpretation" to reflect the establishment of a particular relationship between a part, the present sensible experience, and a whole, the extended temporal history of the agent as subject [Borrett and Kwan, 2008]. As is the case in our own phenomenology in which the interpretation of experience can influence behavior, the establishment of a particular relationship in a robot between its present sensible experience and its extended temporal history translates into a mechanism by which interpretation can modify behavior. Interpretation is distinguished from explanation. Interpretation mediates a subject with an object whereas explanation mediates an object with a framework [Edwards, 1992]. The LIDA model explains the functioning of the robot based on the global workspace theory framework. This explanation is meaningful only to us who interpret these notions within our own horizons. The LIDA model does not present a mechanism by which the robot itself can interpret its experiences, where interpretation here draws on the relationship between the subject and its object both of which occur within its own experience. Although a robot could also possess explanatory mechanisms in its experience, these explanations will only have meaning to the robot itself if they are imbedded or derivative from the level in which interpretation occurs.

1.2. *Synthetic phenomenology*

The issue of whether a physical agent is capable of phenomenal states is a key question in machine consciousness [Gamez, 2009]. Chrisley has introduced the concept of synthetic phenomenology to describe a project that either characterizes the phenomenal states possessed by an artefact such as a robot or uses an artefact, such as a robot, to help specify or model the experience of a subject [Chrisley, 2009].

We have suggested using the form of our experience to define the form of a robot's phenomenal experience to assist in the development agents that are able to interpret their own experiences similar to the way that we interpret our experiences. Because a robot's body differs from our body, it can be argued that the meaning of the experiences in a robot must necessarily differ from ours and that any approach that attempts to extrapolate our phenomenal properties onto a robot's experience represents an exercise in futility. Dreyfus has addressed this dilemma and has expressed pessimism in its resolution [Dreyfus, 2007]. However, Chrisley has made the point that a robot does not have to be conscious to be relevant to the project of synthetic phenomenology [Chrisley, 2009]. Manzotti and Tagliasco made the same point in their discussion of motivation based robotics [Manzotti and Tagliasco, 2005]. We agree with these authors. Dreyfus' concern that we will never know what it is like to be a robot or *vice versa* is relevant only if the robot is conscious. What is sought in the present case are isomorphisms between the form of our own conscious experience and the form of experience, as defined, of a robot. The characterization and modeling can proceed without any ontological commitments concerning conscious experience. There are fundamental aspects features of the form of our experience based on phenomenological analysis such as "presence in absence". The mapping of these aspects onto an autonomous agent then, simply requires that the internal structure and dynamics of the agent have these same features. It will not be necessary to understand what it means to be a robot but it will be necessary for the robot to instantiate a particular structure in its cognitive development defined by this dynamical equivalence.

Having suggested the fundamental form of meaningful experience in a robot to develop this synthetic phenomenology, it is necessary to validate it. This can be accomplished in two ways. If agents with this particular architecture result in behavioral capabilities that are clearly superior to agents that do not have this architecture, then such an architecture becomes functionally validated. The development of such agents represents a long-term goal and the functional validation cannot be expected to be accomplished in the near future. However, maintenance of phenomenological accuracy can be another means of validation. The chosen fundamental form of experience in a robot was proposed based on phenomenological analysis of our own experience. If it can be shown that other aspects of our experience evident on phenomenological analysis arise spontaneously in a robot with such an architecture, then the definition becomes phenomenologically validated. In this regard, it will be shown that the experience of an agent that has access to its temporal horizons is embodied, contextual and normative when applied to EAAs, as is necessary to maintain phenomenological accuracy.

1.3. *Validation of the form of experience in evolutionary autonomous agents*

Because the time scales that are fed back into the robotic controller to instantiate "presence in absence" are continuously available to the agent as it interacts with the environment, they are incorporated into the agent's motor intentionality. Although not sensible, they are what make the sensory motor interaction meaningful from the agent's perspective. If I see a tree branch lying on a path during a stroll through the woods, the branch has a particular significance. If I see the same branch during a stroll through the woods in search of firewood, the situation is experienced differently even though the sensible appearance of the branch on the path is the same. It is being suggested that the difference in these experiences relates to differences in the time horizons which frame the sensible; in the case of looking for firewood, the past requirement for the wood and the future use of the wood is implicit in the experience of seeing the sensible tree branch on the path. In the case of a robot, the difference in experiences of the robot, as the subject of experience, is related to differences in the nature and distribution of time scales to which the robot controller has access. To underline the equal importance of the sensible present experience with the non-sensible time horizons in motor intentionality, that is, to remove the sensible present from its position of priority in neural computation, these components have simply been referred to as the visible and invisible [Merleau-Ponty, 1977]. Because the time horizons in us or the time scales in robots that frame the present sensible experience are available for computation at all times, there is no dissociation between meaning and movement. There is no inner homunculus that assigns meaning to collated sensory data nor is meaning the content of a dispositional attitude that directs movement. With this formalism, the robot's experiences are embodied and its sensory motor interactions with the world are always "pregnant with meaning" [Dillon, 1998].

Perceptual experience is phenomenologically normative. We interact with the environment not to describe it but to get an optimal grip on it [Dreyfus, 2005; Merleau-Ponty, 1962]. Normativity is also a fundamental feature of evolutionary autonomous agents in that optimization based on a given fitness function is the mechanism by which appropriate EAAs are selected. In addition, it is a particular type of behavior that satisfies the fitness function, not an object *per se*. Since it is a behavior that results in optimization, the entire perceptual field can be incorporated into the criteria by which an appropriate EAA is chosen in the evolution of a population. In this way, maintenance of context is also a characteristic feature of EAA development. Care must be taken to insure that the evolutionary procedure that optimizes a robotic interaction with the environment does not present objects to the EAA determinately and out of context such as may occur if only one object is in the arena. Although it is still a behavior that determines fitness in this case, the presentation of an object determinately to the robot removes the richness of experience that is characteristic of our own contextual experience and may prevent the emergence of mechanisms by which context is incorporated. Maintenance of context in the EAA paradigm can thus be assured subject to two further constraints. First, the environment in which a robot evolves must be complicated enough so that context cannot be overlooked. Second, the fitness must be need-based rather than object-based. Biological needs, such as hunger, are not object-based but their satisfaction is the mechanism by which objects acquire contextual meaning. Similarly, a need can be defined in robots such that a particular behavior results in the resolution of the need. In this way, objects that are in the arena are never presented determinately but acquire their contextual meaning as objects as the robots moves in the arena in an attempt to optimize its need-based fitness.

The chosen form of experience thus results in an experience in an EAA that is phenomenologically accurate; its experience is embodied, normative and contextual. The final requirement for validating the implementation of "presence in absence" as the basis for meaningful experience in a robot is to demonstrate how information concerning the agent's entire past history and its future expectations is available through access to its THO. The self-similarity of fractal structures may be a means by which a statistical estimate of the information concerning the history and expectations of an agent can be provided. The dynamics of the neural network controller of properly functioning EAAs has been shown to be scale-free, that is, the dynamics of the EAA controller does not have a dominant or characteristic time scale in the frequency distribution [Borrett *et al.*, 2006]. Power law distributions exhibit long-range correlation or memory and a particular mathematical structure that implements power law distribution is a fractal structure. Fractal structures exhibit self-similar scale invariant properties so that its structure in a short time scale, say, is statistically similar to the structure over a much longer time scale. This property provides a mechanism by which the dynamics of a robot controller over a short period of time can provide statistical information concerning the history of the agent over a much longer time span. We have suggested that the time scales in the frequency decomposition can be a surrogate for the actual time scales in the signal itself. Whether the information concerning the whole temporal history of the agent can be obtained in a manner that confers additional computation abilities to the agent is an empirical question and requires much more work before a final verdict is obtained. In fact, the answer to this one question will determine the fruitfulness of the entire proposed project. However, from a conceptual standpoint, this type of approach to instantiating "presence in absence" is reasonable.

1.4. *Hegel's phenomenology*

Hegel takes as given a particular bimodal structure in consciousness. "But the distinction between the in-itself and knowledge is already present in the very fact that consciousness knows an object at all. Something is for it the in-itself; and knowledge, or the being of the object for consciousness is, for it, another moment. Upon this distinction, which is present as a fact, the examination rests" [Hegel, 1977]. Hegelian phenomenology is simply consciousness watching itself as the relationship between two moments, the object or the in-itself, and our knowledge of the object or the for-consciousness, changes over time; "…since the Notion and object, the criterion and what is to be tested, are present in consciousness itself, …we are spared the trouble of comparing the two and really testing them, so that, since what consciousness examines is its ownself, all that is left for us to do is simply look on" [Hegel, 1977]. If the comparison between the object and knowledge of the object shows that they do not match, consciousness must alter its knowledge to conform to the object. But in altering its knowledge, a new object emerges that is the synthesis of the previous two moments, or "an in-itself-for consciousness". Consciousness now has a new object before it, this "an in-itself-for consciousness", which resulted from "the dialectical movement which consciousness exercises on itself". The science of the experience of consciousness is the description of the flow of these two moments coalescing into a synthesis to produce a new unity. This unity becomes a new object for consciousness and the process is repeated. This dialectic in consciousness continues until knowledge of the object and the object itself, the criteria and what is to be tested, knowledge claims and how they play out in reality, become identical, a state Hegel referred to as absolute knowledge.

For Hegel, dialectic is the movement within consciousness (Taylor, 1975). It is the underlying mechanism by which experience is organized and is distinguished from the knowledge claims within our experience which are the result of this dialectical operation. It is because dialectic can be viewed as mechanism and not simply a descriptive elaboration of phenomenal experience that it is relevant to robotic design. For a robot to develop its own conceptual framework independent of outside programming, it also requires a mechanism by which its experience is organized. Objectively, this mechanism in EAAs will depend on the chosen paradigm and genetic algorithm that constrain the behavioral development of the agents. But if a dialectic represents a fundamental mechanism by which our experience is organized, it is argued that a robot that strives to reproduce our cognitive capacities will also need to reproduce the dialectic within its own phenomenology.

The goal of the experimental section of this paper is modest — it will demonstrate that it is possible for a physical system to instantiate a dialectic within its experience. The paradigm of two moments coalescing into a synthetic unity that is unstable and breaks down into two new moments that becomes the new substrate for the continued dialectical movement of consciousness is the crux of Hegelian phenomenology that will serve as a constraint in the development of cognitive robotics. While the robots interact with the environment, their experiences must also demonstrate these features. Continued development of more sophisticated agents will come from the subset of agents that not only behave optimally but also reproduce this phenomenology.

1.5. *Origin of the dialectic with breakdown*

Morris [2002] has previously discussed a basic problem that arises in any dynamical approach to cognition in which body-world interaction is fundamental. The problem is how is it possible in our phenomenal experience that we can attribute a property to an object rather than associate the property with the body and thing as one moving couple. He used as an example wielding a tennis racket and the experience of the length of the tennis racket. "When I smoothly swing for the ball, I emphasize the resonant modality of body-racket, and what is compelling is an experience of how far the body-racket couple can reach as an unitary system" [Morris, 2002]. But this experience is disrupted if "something intrudes into the body-racket resonance from the outside, as when I hit the ball off-center: Instead of feeling the body and racket resonating as a living couple, I experience the racket as a dead thing reverberating somewhat painfully in my hand, and in that experience I am compellingly aware of touching its mass and length" [Morris, 2002]. With breakdown the experience of the initial body-racket couple during skillful coping changes so that the racket is experienced more determinately and knowledge claims concerning its properties, such as its mass or length, can be explicitly formulated.

With the conceptualization of meaning in a robot as previously defined, it will be shown that the THO associated with skillful coping is scale-free containing no distinct separation of subject and object whereas with breakdown, a bimodal distribution appears. Breakdown will be the mechanism by which the uncoupling of the agent-object experience occurs and the uncoupling is manifest in the bimodal (double-peaked) distribution of the THO. A sensible experience that is associated with a bimodal THO corresponds to an experience that consists of an "in-itself" and a "for-consciousness" whereas a sensible experience that is associated with a scale-free THO corresponds to the experience of skillful coping in which there is no clear separation between subject and object. With the bimodal THO, there is a discrepancy between the object and knowledge of the object whereas with a scale-free THO, such discrepancy has been eliminated. The dialectic in experience emerges as the bimodal distribution reverts to a scale-free distribution with resolution of breakdown only to await the next unexpected perturbation that causes another breakdown. It is feedback of the THO back into the robot controller that provides the robot with the needed information to determine if it is functioning well or if there in infact a discrepancy between the object and knowledge of the object that requires resolution.

The demonstration of the difference in THO distributions will acquire more validity if the sensible experience does not change between skillful coping and breakdown. In Morris' example of hitting the tennis ball off-center, nothing necessarily changes in the sensory data but the consequence of this event is an alteration in experience in which the racket is experienced as distinct from the subject. In the case of an EAA, if the robot's sensible experience changes with breakdown, it could always be argued that any change in the distribution of the THO is related to the differences in sensible experience. Our previous study induced breakdown by placing an obstacle in the arena requiring the robot to move around the obstacle to reach its goal [Borrett *et al.*, 2006]. Clearly, the sensible input to the robot was different between the skillful coping situation and the one associated with breakdown. In the present study, breakdown will be induced not by changing the nature of the arena and the sensible input but by altering the neural network of the robot by lesioning the network. Robots that can still accomplish the task despite the presence of breakdown will be analyzed. In these robots, despite an unaltered sensible experience, the meanings of the experiences will be shown to differ.

2. Methodology

A simulation platform, WEBOTS 3 (Cyberobotics, Switzerland), was used to simulate the movement of a Khepera 2 robot (K-team Corporation, Switzerland). The Khepera is a two-wheeled robot with eight ambient light sensors, eight infrared proximity sensors and a rotation encoder for each wheel. In addition, visual input was derived from a vision module composed of a linear array of 64 photodetectors.

The robot was evolved to explore an arena in which green cylinders and orange blocks were placed. The robot was to spend time around all the cylinders but the blocks had no relevance. Even though the robot was to spend time around the cylinders, the cylinder was never presented determinately such as would occur with a paradigm in which the robot had to identify a cylinder from other different shape objects. The objects, in this case, are simply in the environment.

2.1. *Neural network*

The neural network was composed of 34 fully connected sigmoidal function units. Designated sensory neurons in the network received direct sensory inputs from the sensors of the robot. There were eight proximity, two light, 16 visual and two proprioceptive sensory neurons. The output of the network was determined by two neurons that projected to the wheels. The remaining four neurons were interneurons; that is, they neither received direct sensory inputs nor projected to the wheels. The neurons in the network were updated synchronously. Activity of one interneuron was chosen as reflective of the network dynamics.

2.2. *Adaptive synapses and Hebbian adaptation rules*

The algorithm used in the evolutionary process involved genetically encoding a set of local adaptation rules for each node in the network rather than the traditional evolution of fixed synaptic weights. The approach used is a modification of that described by Urzelai and Floreano [2001]. As the robot interacted with the environment, the synaptic weights were subjected to long-term modulation by the activities of presynaptic and postsynaptic neurons (Eq. (2.1)).

$$w_{ji}^{t} = w_{ji}^{t-1} + \eta_{ji} u_{ji}, \tag{2.1}$$

$$u_{ji} = h(y_j, y_i, w_{ji}). \tag{2.2}$$

The synaptic weight ω_{ji}^{t} for the sensory-motor cycle τ is updated according to the synaptic weight of the previous cycle ω_{ji}^{t-1}, the synaptic adaptation rate η_{ji}, and the weight update υ_{ji}. This value is a function of the presynaptic activity y_i, the postsynaptic activity y_j, and the current synaptic weight ω_{ji} (Eq. (2.2)).

Hebbian adaptation rules (Eqs. (2.3)−(2.7)) were used to determine how these parameters influence υ_{ji}.

$$h_1 = (1 - w_{ji}) y_j y_i, \tag{2.3}$$

$$h_2 = w_{ji}(y_i - 1) y_j + (1 - w_{ji}) y_j y_i, \tag{2.4}$$

$$h_3 = w_{ji} y_i (y_j - 1) + (1 - w_{ji}) y_j y_i, \tag{2.5}$$

$$h_4 = \begin{cases} (1 - w_{ji}) d(y_i, y_j), & d(y_i, y_j) > 0, \\ w_{ji} d(y_i, y_j), & d(y_i, y_j) > 0, \end{cases} \tag{2.6}$$

$$d(y_i, y_j) = \tanh(4(1 - |y_i - y_j|) - 2). \tag{2.7}$$

Briefly, the weight of a synapse governed by the plain Hebbian rule η_1 is strengthened proportionally to the activation of presynaptic and postsynaptic neurons. The postsynaptic rule η_2 is similar to the plain Hebbian rule, but it also weakens the synapse when the postsynaptic neuron is active while the presynaptic is not. The presynaptic rule η_3 is the converse of η_2. The covariance rule η_4 strengthens the synapse when the postsynaptic and presynaptic activities are similar, and weakens it otherwise.

Additionally, a fifth rule η_5 was included to allow some synapses to have fixed weights that do not change during the life of a robot.

$$h_5 = 0. \tag{2.8}$$

2.3. *The genetic algorithm*

The properties of the neural network described above were encoded on the robots' artificial chromosomes and subjected to evolution by the genetic algorithm. Each chromosome had 34 genes corresponding to each node in the network. Each gene had four sequential components and all components used real-value encoding; the first component was the initial 34 synaptic weights for the 34 outgoing projections from the node, the second the adaptation rate, the third the polarity of the outgoing signal and the fourth the Hebbian rule.

The experimental setup is shown in Fig. 1. A population of 100 robots was initialized with each robot having a random set of genes. At birth, the activities of the neurons are randomly initialized in the range [0, 1], and the weights are set to the inherited initial weights ϖ_{ji}. A single robot was placed randomly in any of the four corners of the arena and was oriented toward the center. The life of each robot consisted of 1500 sensory-motor cycles, each lasting 100 ms. The performance of each robot was assessed by a fitness function based on two measures of exploration (see below). At the end of each generation, a new population of 100 individuals was

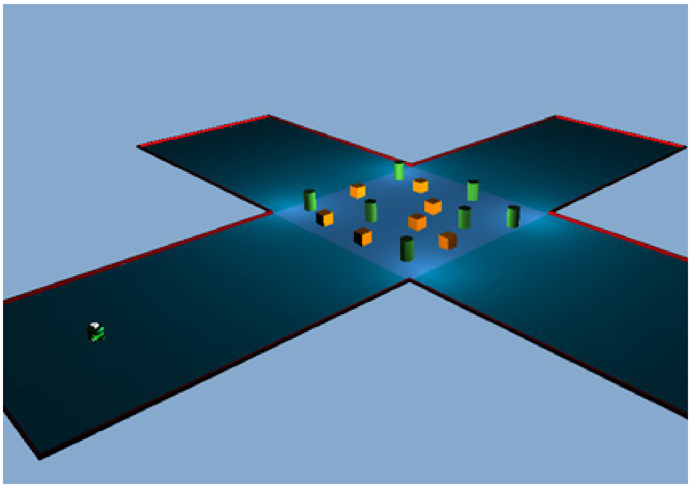

Fig. 1. A view of the arena showing the robot in the lower left corner approaching the cylinders (targets) interspersed among the blocks (obstacles) in the center of the arena.

produced by a combination of elitism, sexual reproduction, and asexual reproduction. The best 20 individuals were cloned without genetic modifications. The remaining progeny were produced by fitness-proportional roulette wheel selection; 90% of robots mated sexually and crossover between chromosomes occurred at two random locations. Genetic loci were subsequently mutated at a mutational probability of 2%, which was reduced to 0.2% after the robots achieved fitness. Populations were evolved until a plateau in fitness occurred. This usually took 300 generations.

2.4. *The fitness function*

The fitness function employed two measures of exploration. The first measure assessed the movement trajectory (Eq. (2.9)).

$$\Delta f_1 = \alpha(o_1 o_2)^2 + \beta(1 - |o_1 - o_2|)^2. \tag{2.9}$$

The fitness update Δf_1 is based on the output activities o_1 and o_2 of the two motoneurons; the factors α and β determine the relative influence of each term of the equation. The first term determined the distance covered by the robot and the second the linearity of the movement trajectory. The latter constraint was needed to prevent the robots from spinning in circles. Thus, robots that covered more distance and travel in straighter paths received more fitness points.

The second fitness measure was proportional to the number of sensory-motor cycles that the robot spent nearby on a cylinder. The fitness points awarded each cycle nearby a cylinder diminished linearly over time (Eq. (2.10)) in order to force the robot to approach multiple cylinders.

$$\Delta f_2 = \mu t + \kappa. \tag{2.10}$$

The fitness update Δf_2 is given by Eq. (2.10) if the robot is within a defined radius of a cylinder and is set to zero otherwise. The time spent nearby on a cylinder, t, was independently monitored for each cylinder. The parameters μ and κ determines the decay rate and the maximum award, respectively.

2.5. *Lesion analysis*

Two populations were evolved to explore the arena by moving around cylinders and avoiding squares. The interneurons were lesioned by setting the weights of the synapses between the interneuron and all the neurons of the network, including all incoming and outgoing connections, to zero. Different generations in the two populations were sampled to find agents that were still able to perform the task despite the presence of the lesion. Once an agent was found that still could accomplish the task despite the lesion, a comparison of the THOs associated with the experience of the robot before lesioning and after lesioning was made. The THO was calculated through a Fourier analysis of the time series for the entire 1500 sensory-motor cycles of a remaining interneuron. Statistical significance was confirmed by the application of the Kolmogorov−Smirnov goodness-of-fit test (K−S test).

3. Results

Agents were found in the first population that were still able to perform the task despite the presence of lesions. In the second population, no such agents were found. In the first population, five generations were identified in which there were agents that could successfully accomplish the task despite the presence of lesions. Five agents from each of these five generations were identified and the task was performed before the lesion and after the lesion for each of the agents. The time series from the interneuron was used in the calculations. The single trial data from these five agents in one generation were averaged to give a final distribution for that generation. This was done for each generation yielding five prelesion and postlesion distributions.

The data for one generation is shown in Fig. 2. The power spectral density plot reflects the degree to which a particular time scale contributes to the overall signal. The time scales were identified with specific frequencies in the Fourier analysis and the power of these frequencies is displayed. Before the lesion, the distribution is scale-free in that there is a gradual decrease in power from low frequencies to higher frequencies that follows a power law relationship $(1/f)$. After the lesion, a peak occurs at the higher frequencies and is maximal at arbitrary frequency unit 300. This gives the distribution a bimodal appearance with one peak at very low frequencies and the second peak at higher frequencies. The changes in Fig. 2 was prototypical for all five generations with all five generations showing a scale-free distribution before the lesion and a bimodal distribution after the lesion. The K−S test was statistically significant for all five generations.

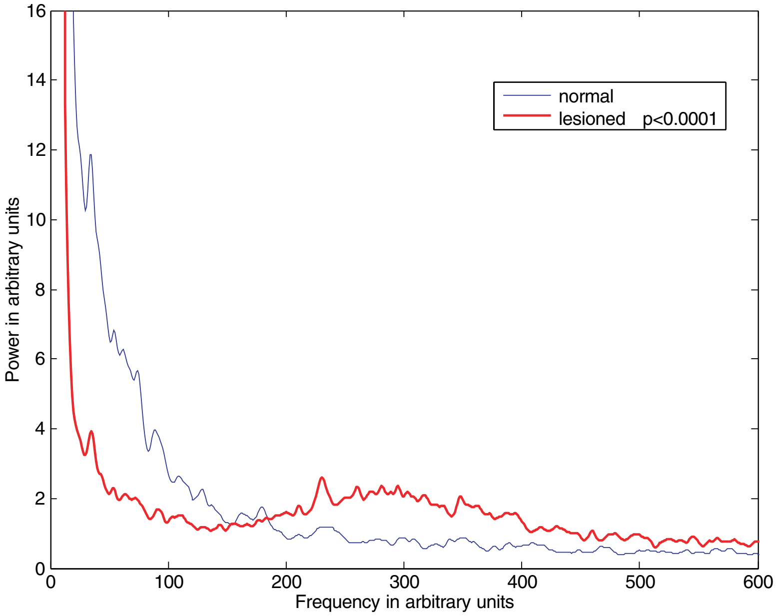

Fig. 2. Power spectral density plot (THO) of the activity of a hidden unit before and after network lesion. The THO prior to lesioning is scale-free whereas the THO after lesioning has a bimodal distribution with one peak as the frequency approaches 0 and the second peak at about 300 units.

4. Discussion

The goal of this paper was modest — to propose a regulative definition of meaningful experience in an autonomous agent and, with this definition, to show how a dialectic could emerge in such an experience. By taking the fundamental form of experience as instantiating "presence-in-absence" and using evolutionary autonomous agent algorithms to maintain embodiment, contextuality and normativity, the approach demonstrated that an autonomous agent's experience can alternate between two poles. During skillful coping, the agent's sensible present is framed by a THO that has a $1/f$ distribution whereas with breakdown, the agent's THO can have a bimodal distribution. This latter distribution corresponds to the Hegelian experience of the "in-itself" and the "for consciousness" whereas the former corresponds to the Hegelian experience of a synthesis of these moments. A dialectic in experience emerges as the bimodal distribution reverts to a scale-free distribution with resolution of breakdown only to await the next unexpected perturbation that causes another breakdown. It is suggested that through repeated resolution of breakdown, those agents in which phenomenal experience is characterized by a dialectic will eventually possess the mechanism by which a world of enduring objects will emerge in that experience. Once established, agents with this mechanism can then be used for specific tasks and will be able to accomplish these tasks through a mechanism that has not been prestructured by our interpretive frameworks.

It has been assumed that empirical reality, our experience of being a distinct entity in an objective world of enduring objects, is synthesized by the nervous system. The mechanism of synthesis and the outcome of the synthesis, which is our empirical experience, are distinct. In the Kantian synthesis to which Hegel was responding, the *a priori* categories shape the spatiotemporal manifold in intuition to result in the contents associated with empirical reality. In Hegel's synthesis, a dialectic that emerges out of our interaction with the world structures experience. Because it is interaction with the world that shapes the experience of an empirical reality we can say that Hegel grounds the possibility of experiencing the world, not in *a priori* categories, but in the world itself. When this conceptualization is translated into robot design, we can also say that the dialectic in the robotic experience ensures that any discrepancy between the object and knowledge of the object always results in a new dynamics in which this discrepancy has been eliminated and the robot can function skillfully. The requirement of the presence of a dialectic in the robot's experience as a constraint during evolution could therefore provide a means by which the need for external supervision during learning is eliminated.

This approach takes robotic design to consist of two sequential parts. The first part is the evolution of robots that incorporate a mechanism in which interpretation of their own experience is possible. The second part consists of evolving the agents to accomplish a specific task. The distinction between evolution and learning has been well described by Nolfi *et al.* [1994]; evolution is a process of selection of a population of individuals whereas learning relates to modifications within an individual during its own lifetime. They make the point that agents can learn tasks that differ from the tasks for which they were evolved, a characteristic required if the agent is to possess cognitive capabilities similar to our own. Manzotti and Tagliasco made the same point in their motivation-based robots in which they explicitly separated the ontogentic part from the phylogenetic part in order to demonstrate how a robot could develop an "interest" in an activity even though phylogentically, this interest was not present at the time of inception [Manzotti and Tagliasco, 2005]. In our case, agents are evolved phylogenetically to experience a dialectic as they interact with the environment. Learning a particular task ontogenetically would then incorporate the same mechanism even though the learned task was not used in the initial evolutionary phase.

Although meaningful experience has been identified with neural activity within the neural controller of the agent, this does not represent a return to the notion of a transcendent subject processing information from an external world. The formalism remains true to phenomenological analysis. The understanding that was modeled puts understanding within movement and not prior to it as would be assumed with the model of a transcendent subject independently controlling its own movement within the world or after it in the case of a transcendent subject collating sensory data. Both in the case of skillful coping in which there is no clear distinction between subject and object within experience or in the case of breakdown in which the sensible experience is framed by a bimodal invisible, there is no separation between meaning and movement. Even though the meaningful experience arises within the agent's controller while the agent is interacting with an external world, the movements themselves are meaningful.

The paradigm used to induce breakdown adds more strength to the phenomenological argument. We previously have demonstrated in an EAA evolved to move to a location in an arena using a two-step movement that the introduction of the obstacle into the arena resulted in a change in the time scale distribution with an increase in short-term time scales [Borrett *et al.*, 2006]. Two problems existed in the interpretation of this data. In the first place, the robot had to move to only one location in the arena. The location, therefore, was presented determinately since context was not relevant in the behavior of the robot. This lack of phenomenological accuracy in the experimental paradigm makes the results suspect. Secondly, the adoption of a reactive strategy from an automatic strategy in moving around the obstacle could raise the concern that the change in THO simply reflected this reactive motor strategy. In this paper, the confounding effect of a change in the perceptual field or a change in motor strategy in the interpretation of the change in the THO is addressed. By introducing breakdown through a lesion in the robot's network and by choosing only those robots that could still perform the task, we have eliminated these confounding factors. Since breakdown results in the establishment of a bimodal distribution in the THO from a scale-free distribution, the meaning of the experience to the robot changes even though its sensory input and behavior are unaltered.

Many issues remain and multiple directions can be adopted. All that was demonstrated experimentally is that it is possible to find robots with the requisite bimodal nature of experience to support the arguments presented concerning the relationship between Hegelian phenomenology and robotics. Elaboration of the argument will require extensive evolution of these identified robots to determine if further structures emerge that remain consistent with phenomenological analysis. In addition, the benefit that phenomenological accuracy as a constraint confers over and above a robotic architecture in which the phenomenology is irrelevant needs to be demonstrated. For Hegel, dialectic is the movement within our conscious experience. By insisting that an agent has the same movement within its experience while it interacts with the environment, an approach is suggested that may facilitate the development of cognitive agents that can function independently.

Acknowledgments

This study was funded by a research grant from the Toronto East General Research Foundation. We would like to thank David Morris for his comments during the preparation of the manuscript.

References

Baars, B. and Franklin, S. [2009] "Consciousness is computational: The LIDA model of global workspace theory," *International Journal of Machine Consciousness* **1**(1), 23–32.

Borrett, D. S., Khan, S., Lam, C., Li, D., Nguyen, H. and Kwan, H. C. [2006] "Evolutionary autonomous agents and the naturalization of phenomenology," *Phenomenology and the Cognitive Sciences* **5**, 351–363.

Borrett, D. S. and Kwan, H. C. [2008] "Understanding understanding: An evolutionary autonomous agent approach," *Review of General Psychology* **12**(2), 137–146.

Chemero, A. and Silberstein, M. [2007] "After the philosophy of mind," *Philosophy of Science* **75**, 1–17.

Chrisley, R. [2009] "Synthetic phenomenology," *International Journal of Machine Consciousness* **1**(1), 53–70.

Dillon, M. [1998] *Merleau-Ponty's Ontology* (Northwestern University Press, Evanston, IL), p. 212.

Di Paolo, E. [2002] "Organismically-inspired robotics: Homeostatic adaptation and teleology beyond the closed sensory-motor loop," in *Dynamical Systems Approach to Embodiment and Sociality*, Murase, K. and Asakur, T. (eds.) (Advanced Knowledge International, Adelaide, Australia), pp. 19–42.

Dreyfus, H. [2005] "Merleau-Ponty and recent cognitive science," in *Cambridge Companion to Merleau-Ponty* (Cambridge University Press, Cambridge, UK).

Dreyfus, H. [2007] "Why Heideggerian AI failed and how fixing it would require making it more Heideggerian," *Artificial Intelligence* **171**, 1137–1160.

Edwards, T. [1992] "Interpretation, explanation and culture," *Crit Quart* **34**(3), 96–106.

Floreano, D. and Mondada, F. [1996] "Evolution of homing navigation in a real mobile robot," *IEEE Transactions on Systems, Man, and Cybernetics B* **26**(3), 396–407.

Gadamer, H. [2000] *Truth and Method* (Continuum, New York), p. 304.

Gamez, D. [2009] "The potential for consciousness of artificial systems," *International Journal of Machine Consciousness* **1**(1), 213–223.

Hegel, G. W. F. [1977] *Phenomenology of Spirit* (Oxford University Press, Oxford).

Manzotti, R. and Tagliasco, V. [2005] "From behavior-based robots to motivation-based robotics," *Robotics and Autonomous Systems* **51**, 175–190.

Merleau-Ponty, M. [1962] *Phenomenology of Perception* (Routledge and Keegan Paul, London).

Merleau-Ponty, M. [1968] *The Visible and Invisible* (Northwestern University Press, Evanston, IL).

Morris, D. [2002] "Touching intelligence," *Journal of Philosophy Sports* **29**, 149–162.

Nolfi, S., Elman, J. and Parisi, D. [1994] "Learning and evolution in neural networks," *Adaptive Behavior* **3**(1), 5–28.

Nolfi, S. and Floreano, D. [2000] *Evolutionary Robotics* (MIT Press, Cambridge).

Russon, J. [2003] *Human Experience Philosophy, Neurosis and the Elements of Everyday Life* (State University of New York Press, Albany, NY).

Sun, R. [2007] "The importance of cognitive architecture: An analysis based on Clarion," *Journal of Experimental and Theoretical Artificial Intelligence* **19**(2), 159–193.

Taylor, C. [1975] *Hegel* (Cambridge University Press, Cambridge), p. 129.

Urzelai, J. and Floreano, D. [2001] "Evolution of adaptive synapses: Robots with fast adaptive behavior in new environments," *Evolutionary Computation* **9**(4), 495–524.