

I. Intellectual Merit

1. Background and Significance

Animals exhibit repertoires of behavior that are remarkably rich, with both continuous modulation of ongoing behavior and discrete switches between different behaviors as changing circumstances warrant [26, 58, 84]. The flexibility and robustness of the behavior of even simple animals puts our most sophisticated technological achievements to shame, and we would very much like to harness these capabilities in the design of future artificial systems [63, 85, 116]. Doing so, however, requires that we first understand the mechanisms of behavioral modulation and switching in biological systems.

How is it that a single nervous system can generate such a diversity of behavior? For many years, the dominant explanation was that distinct neural circuits underlie different behaviors, with specialized neural circuitry such as command neurons responsible for switching between them [49, 87]. However, decades of subsequent work has demonstrated the distributed and dynamical nature of neural function and the fact that any given neuron, circuit or brain region can participate in many different behaviors. This capability has been referred to as polymorphism [56], reorganization [99], multifunctionality [86, 145], reuse [4], network flexibility [7, 8], the dynamome [83], or the chronnectome [29-30]. We will use the term *multifunctional* in this proposal. In a multifunctional circuit, neural activity is viewed as a dynamic property that can be modulated and reshaped as necessary rather than as something fixed by its synaptic connectivity [25]. Proposed mechanisms of reconfiguration include external inputs (e.g., sensory receptors or projection neurons from other circuits modifying network dynamics by selectively exciting or inhibiting specific components of a target circuit) [9, 22, 98] and neuromodulation (which can selectively modulate neural excitability or synaptic efficacy) [43-44, 91]. In all cases, a multifunctional neural circuit is essentially viewed as an autonomous dynamical system from which extrinsic influences can select a particular attractor.

This way of thinking about the role of multifunctional neural circuits in behavioral switching makes a critical assumption: like a puppeteer pulling the strings of the body as it dances upon the stage of the environment, the nervous system alone is responsible for behavior. Increasingly, this assumption has been called into question by work on embodiment and situatedness in neuroscience, cognitive science, robotics, and philosophy of mind [33, 36, 79, 102, 109, 123, 134]. Broadly speaking, situatedness concerns the role played in an agent's behavior by its ongoing interactions with its immediate environment. Embodiment, in contrast, concerns the role of the structure and properties of an agent's body in its behavior. From an embodied and situated perspective, a nervous system is but one component of a larger brain-body-environment (BBE) system (Figure 1) and the mechanisms of behavior are to be found not in

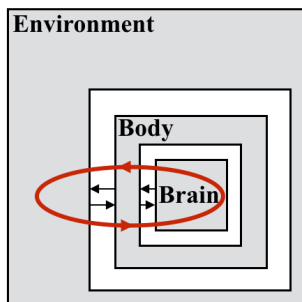


Figure 1: Behavior (red) arises from the interaction between brain, body and environment.

any one component, but rather in the dynamics of their interaction [10]. How else to explain, for example, the fact that the same neural activity can produce different behaviors in different circumstances due to the context-sensitivity of the consequences of muscle activation (a kind of biomechanical multifunctionality) [101]? Or that different gaits can arise from the same pattern generation circuit depending on differing sensory feedback from the different media through which an animal moves [24, 133]? *In a brain-body-environment system, stable behavior does not in general correspond to simply an attractor of a neural circuit, but rather to an attractor of the entire BBE system in which that neural circuit participates.* Said in a slightly more technical way, it is the nonautonomous dynamics of a

neural circuit when embedded in the flow of feedback through the body and environment that is most relevant for understanding the “neural basis” of behavior, not its autonomous dynamics in isolation from that feedback. After all, brains, bodies and environments co-evolve and it is the entire package that is selected for survival and reproductive success.

How does an embodied and situated perspective impact the way we should think about behavioral switching in general and the role of neural multifunctionality in particular? If we view behavior as an attractor of the brain-body-environment system, then behavioral switching involves the destabilization of an existing BBE attractor and the stabilization of a new one. Such a switch can be initiated by changes in any of the components and involves the consequences of

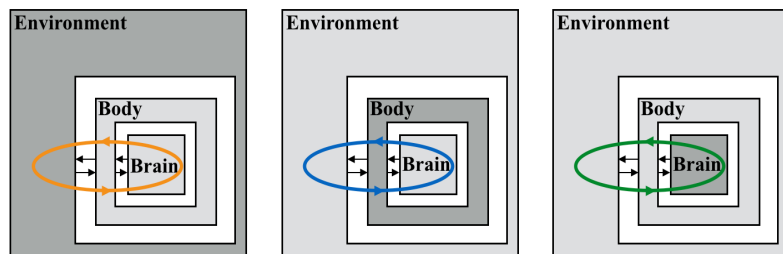


Figure 2: A change in the dynamics of any component (dark gray) can trigger a switch to a new behavior of the entire BBE system (different colors), thereby exercising different dynamical regimes of all components.

that change playing through the entire system (Figure 2). For example, a change in the viscosity of the medium through which locomotion occurs modifies the movements that a given pattern of muscle contractions produce, which in turn changes the pattern of sensory feedback that the underlying neural circuit receives, potentially shifting it into a new pattern of activity that ultimately manifests as a different gait [24]. Thus, viewed separately, all compo-

nents of a BBE system can exhibit multifunctionality in the sense that different patterns of coordination can be observed within them during different behaviors. We will refer to multifunctionality within a BBE system as *embodied multifunctionality*. Within this context, questions about neural multifunctionality become questions about what kind of nonautonomous dynamics can be evoked from an anatomically fixed neural circuit when it is embedded in a given sensorimotor feedback loop through its body and environment.

Unfortunately, the BBE perspective raises severe experimental challenges. Studying any one component of a brain-body-environment system is difficult enough, but studying all three components and their interactions in any animal is currently beyond our experimental capabilities. Although substantial progress is being made using optogenetic imaging [112], it is not yet feasible to monitor and manipulate all relevant neural activity within the nervous systems of intact, freely-behaving animals, let alone the key biomechanical and environmental properties. More fundamentally, we currently lack the analytical tools necessary to understand large networks of densely-interconnected, heterogeneous, nonlinear, dynamical elements whose interactions span the brain-body and body-environment boundaries. For this reason, we and others use evolutionary algorithms [5, 57, 61, 93] to evolve model BBE systems and then analyze their operation using the mathematical tools of dynamical systems theory and information theory. By mimicking the process by which biological brain-body-environment systems were produced, we can likewise evolve model agents that can fully exploit the freedom to partition solutions across the brain-body-environment boundaries in ways that will not necessarily align with our preconceptions about how such systems should work. Such evolved model BBE systems have the significant advantage that we have complete access to and control over their nervous systems, bodies and environments, allowing us to undertake detailed analyses of their operation.

Our goal is to utilize the evolution and analysis of model brain-body-environment systems to explore the role of embodied multifunctionality in behavioral switching. Specifically, we seek to answer the following questions: (1) What role does embodiment play in evoking different activity patterns from the same neural circuit?, (2) How can we identify and characterize

the distinct information architectures that arise in embodied circuits?, and (3) How does embodied multifunctionality scale with behavioral complexity?

2. Previous and Preliminary Work

We briefly present several lines of previous and preliminary work that support the research described in this proposal. First, we demonstrate that model brain-body-environment systems, including those involving multiple behaviors, can be reliably evolved. Second, we show that such BBE systems can be analyzed dynamically. Finally, we provide evidence that such systems can be fruitfully analyzed information theoretically.

2.1 Model brain-body-environment systems can be evolved

In any given BBE model, three components must be specified: the neural model, the body model and the environment model. In addition, when evolving such a system using an evolutionary algorithm, a fitness measure to be optimized must be provided. The fitness measure, body model and environment model vary greatly from task to task, but the neural model and evolutionary algorithm are generally fixed. As a neural model, we use continuous-time recurrent neural networks (CTRNNs) [11]. CTRNNs are known to be universal dynamics approximators when the parameters are unconstrained [35, 54, 78]. An N -neuron CTRNN has up to $N^2 + 2N$ free parameters (N^2 weights, N time constants and N biases). These parameters are set by a real-valued evolutionary algorithm with Gaussian mutation and modular crossover. Both our CTRNN and evolutionary algorithm are publicly available [15]. We only employ behavioral fitness functions, that is, fitness functions that select for some overall property of the entire brain-body-environment system rather than particular motor neuron output trajectories. This allows us to explore the space of possible solutions to a given task with a minimum of preconceptions about how a given behavior ought be accomplished.

This general approach to the evolution of dynamical BBE models was initiated almost 30 years ago [17] with the more applied aspects of this methodology subsequently christened “evolutionary robotics” [23, 37, 103]. Since then, both we and others have successfully applied it to a very wide range of both animal and robot behaviors, as well as to general theoretical questions in neuroscience and cognitive science. Some examples include locomotion [17, 59, 64, 73, 81, 82, 108, 119], chemotaxis [17, 71], navigation [51, 136], learning [70, 104, 110, 132, 147], categorization [12, 18, 132], communication [6, 45, 94, 115, 143], and social interactions [52, 95, 96]. Two examples are particularly relevant to the work described in this proposal.

In the first example, we evolved a CTRNN capable of producing coordinated walking in a 6-legged body. Each leg was composed of a segment connected to the body by a joint actuated by two opposing “muscles” and a binary (up/down) foot. When the foot was “up” (swing phase), any torque produced by the muscles served to swing the leg along an arc relative to the body, with a maximum angular acceleration and angular limits. When the foot was “down” (stance phase), any torque produced by the muscles applied a translational force to the body under Newtonian mechanics. The leg was only able to generate force over a limited angular range of motion (modeling how mechanical advantage changes with limb geometry). A stancing leg exceeding these limits could still provide support, but only within a wider range of limits (modeling skeletal constraints). When a stancing leg reached these hard kinematic limits, forward motion ceased (modeling a loss of postural stability). When the leg subsequently lifted its foot, it immediately snapped back to the swing angular limits (modeling the passive restoring force of muscle). Each leg also had an angle sensor whose output was proportional to the angular deviation of the leg from perpendicularity to the long axis of the body. Fitness was determined simply by measuring the total forward distance that the body moved in a given period of time. Among other things, this BBE model was used to examine the conditions under which different pattern generator organizations evolve, the dynamical structure of these different organizations,

the impact of network architecture on walking performance and evolvability, the role of neuromechanical interaction in pattern generation, and the application of evolved locomotion controllers to legged robots [10, 14, 16-17, 34, 55, 114]. Variations of this model have also been employed by others (e.g., [60, 69, 81-82, 125]).

In the second example, we evolved agents that could “visually” distinguish between two classes of objects that they encountered by catching objects of the first class (circles in the original experiments) and avoiding objects of a second class (diamonds) [12]. The agent could move back and forth along a horizontal line while objects fell toward it from above. Using an array of seven rays, the agent could sense the distance to each point of intersection between a ray and a falling object. Fitness was measured so as to minimize the final horizontal separation for circles and maximize the final horizontal separation for diamonds. The highest performing agents reliably evolved strategies involving active perception. In a relational extension of this task, we evolved agents to visually discriminate between objects based upon relative size [18]. In a spatial version of this relational categorization task, both objects were presented to the agent at the same time and the agent had to choose the smaller one by catching it as it fell. In a temporal version of this task, a reference object was first presented. After the reference object was removed, a second probe object was dropped and the agent had to catch the probe object if it was smaller than the reference object and avoid it if it was larger. Variations of these tasks have been subsequently employed by many other researchers (e.g., [3, 39, 92, 137, 138]).

More recently, research has slowly begun to move beyond the evolution of single behaviors to the evolution of multiple behaviors in the same brain-body-environment system, although this is still very much a frontier activity in evolutionary robotics. This is the work we plan to directly build upon for our examination of the mechanisms of embodied multifunctionality. When describing any particular behavior, one typically focuses on a narrow range of timescales. But a brain-body-environment system has dynamics on many timescales. As the nervous system, body and/or environment changes on these other timescales, the behavior of interest may be modulated or even switched to a completely different behavior. Thus, it is convenient to classify simple behavior switches by the BBE component whose change triggered the switch (Figure 2). In the first class, changes in the environment trigger a behavioral switch. For example, a variation of the visually-guided agent described above has recently been evolved to switch between catching any circles that it encounters, avoiding any horizontal lines, and balancing any poles [31]. A version of this task will serve as the starting point for Specific Aim 2. In the second class, changes in the body trigger a behavioral switch. For example, a CTRNN was evolved that produced chemotaxis when placed in a chemosensory body and walking when placed in a legged body [69]. This task will serve as the starting point for Specific Aim 1. Other examples of this class include switches triggered by leg growth [17], effector damage [142], metabolic change [2], and visual inversion [46]. In the third class, changes in the nervous system, such as from neuromodulation or synaptic plasticity, trigger a behavioral switch. In our study of embodied multifunctionality in this proposal, we focus on the different neural activity evoked by body- and environment-triggered behavioral switches.

2.2 Model brain-body-environment systems can be analyzed dynamically

Our main interest in evolving BBE models is not in the evolution per se, but in the lessons that studying the operation of the evolved BBE systems have to teach us about the roles of situatedness, embodiment and neural dynamics in behavior. Accordingly, our focus is largely on analysis. One class of mathematical tools that we utilize is dynamical systems theory (DST), including finding limit sets, their stabilities, the phase portraits that organize them, and the bifurcation diagrams and parameter charts that organize those phase portraits [89, 128]. Our basic approach has been the following. First, we characterize the phase portrait of the entire BBE system as an autonomous dynamical system. Second, we decompose the coupled system into non-

autonomous agent and environment subsystems, perform a bifurcation analysis of the autonomous dynamics of each using the coupling inputs as bifurcation parameters, and then study how the time-varying inputs from each component drive the other component across its bifurcations using a quasistatic approximation. Third, we further decompose the agent into nonautonomous body and nervous system subsystems and perform a similar analysis. Finally, we apply a similar analysis to decompose the evolved neural circuits. Our dynamical characterization of the neural circuits is grounded in detailed studies of the general dynamical behavior of CTRNNs [11, 13].

We have applied dynamical analysis to a number of evolved BBE systems, including walking, chemotaxis, learning, object discrimination and relational categorization. For example, in the walking task, we characterized the operation of sensory-driven walkers in terms of bifurcations across a bistable region in the neural dynamics [10], decomposed the operation of network oscillators into dynamical modules [34], and quantitatively accounted for parameter variation among the top performing walkers in terms of multiple degeneracies in the maps from neural parameters to neural dynamics to embodied behavior to fitness [14]. This analysis led to several qualitative predictions about the structure of variability in populations of pattern generators that were subsequently observed in biological pattern generators [16]. It also led to a method for designing multipattern generators which was then implemented in analog VLSI [77]. As a second example, we analyzed the dynamics of the best evolved object discrimination agent at three different levels [12]. A major goal of the present proposal is to extend such dynamical analysis to BBE systems exhibiting switching between multiple behaviors.

2.3 Model brain-body-environment systems can be analyzed information-theoretically

Another class of mathematical tools that we have used to analyze evolved BBE systems is information theory (IT). In addition to standard IT tools, we have also developed new tools as part of an earlier NSF grant to PI Beer (IIS-0916409). A particular focus of this earlier work was building on the traditional information theoretic tools (including entropy and mutual information [38], specific information/surprisal [28, 41] and transfer entropy [121]) to develop a mathematical toolkit for characterizing the dynamics of information flow. A central component of this toolkit was partial information decomposition (PID), a method for decomposing multivariate mutual information into combinations of unique, redundant and synergistic contributions [140]. On this foundation, we also proposed measures of information gain, loss and transfer, the latter correcting a shortcoming in transfer entropy which causes it to conflate unique and synergistic contributions [74, 141]. Collectively, these tools allowed us to track information about an external stimulus as it flows throughout a neural circuit. The further development and application of PID and related methods has become a very active area of research (e.g., [20, 50, 75-76, 90, 130, 139]).

We have applied such information theoretic tools to a number of evolved BBE systems. For example, we performed an information-theoretic analysis of the embodied relational categorization model mentioned above [18]. Using a combination of time-varying mutual information, specific information, and our information gain and transfer measures, we characterized how information about the reference object size flowed through the sensors and into the interneurons. We then examined how this was stored after the reference object had disappeared from the agent's field of view. Finally, we studied how information about the relative size of the two objects flowed through the system once the probe object appeared. Our analysis revealed two different strategies for information storage among the highest performing agents. In the first class, reference object size was stored in some combination of interneuronal state. However, in the second class, we showed that the interneuronal state contained no information about reference object size after that object disappeared from the field of view. Instead, this second class of agents stored this information in the position of its body relative to where the object fell, an ex-

ample of what is sometimes called information offloading. As a second example, we analyzed the flow of information about salt concentration through an embodied model of the *C. elegans* klinotaxis circuit [72]. Novel features of this work included a description of how the locomotion pattern generator gated the flow of concentration information from interneurons to the neck motor neurons controlling steering and the identification of a common information architecture (pattern of information transfer between identified neurons) across the ensemble of evolved klinotaxis models despite significant variation in individual neural and synaptic parameters. A major goal of the present proposal is to extend such information theoretic analysis to behavioral switching in BBE systems.

3. Proposed Work

3.1 What role does embodiment play in evoking different activity patterns from the same neural circuit?

A fundamental challenge raised by embodied multifunctionality is understanding how different activity patterns can be evoked from the same neural circuit as the feedback through the body and environment in which it is embedded changes. To understand the dynamical underpinnings of a behavioral switch, we must (1) characterize the BBE mechanisms by which the original behavior was stabilized, (2) identify the location and nature of the change that initiated the switch, (3) track how the consequences of that change propagate around the BBE feedback loop to destabilize the previous behavior and provide a transient path to the new one, and (4) characterize the BBE mechanisms by which the new behavior is stabilized. We propose to develop the tools necessary to address these challenges using the chemotaxis/walking body switch task described earlier, in which the same neural circuit produces walking behavior when placed in a legged body and chemotactic behavior when placed in a chemosensory body [69]. Of course, neither animals nor robots switch bodies on a regular basis. However, this task captures the effects of a radical change in the feedback through the body and environment on the operation of a neural circuit. It is thus rich enough to fully engage the theoretical question raised in this specific aim but simple enough to be analyzed in some detail.

The model setup is as follows. The legged body consists of a single leg that can swing forward or backward relative to the body with a foot that can grasp and release the substrate (Figure 3, left). The leg is controlled by three effectors: backward swing (BS) and forward swing (FS) muscles and a foot (FT). In addition, there is a leg angle sensor (ϕ). The complete equations of motion for this body model, including position and velocity limits and passive responses, have three degrees of freedom and are given in [14]. The chemotaxis agent has a circular body with

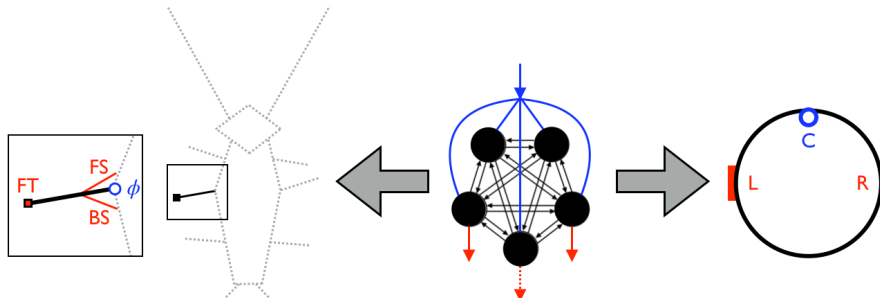


Figure 3: The same neural circuit (center) must generate walking when coupled to a legged body (left) and chemotaxis when coupled to a chemosensory body (right). Sensory input is shown in blue and motor output is shown in red.

two wheels whose speed is controlled by two effectors (L and R) and a single chemosensor C, placed in an environment containing a single chemical source (Figure 3, right). The equations of motion for this body model also have three degrees of freedom and are described in [69]. The

neural circuit is an N -neuron ($N \geq 3$) fully-connected CTRNN with N weights from a sensory input S for a total of $N^2 + 3N$ free parameters. In the legged body, three of these neurons drive the BS , FS and FT effectors and the remaining $N - 3$ are interneurons (Figure 3, middle). In the chemotaxis body, two of these neurons drive the effectors L and R and the remaining $N - 2$ are interneurons. Finally, the sensory input S of the circuit is attached with appropriate scaling to the leg angle sensor ϕ for walking and to the chemosensor C for chemotaxis.

Our first step will be to evolve at least 100 each 3-, 4- and 5-neuron circuits for this task. Evolving this many circuits for each condition ensures that we have a good sample of possible solutions to work with. Based upon our previous experience with a limited number of evolutionary runs, we know that 3-neuron circuits are capable of solving this task, but that 4- and 5-neuron circuits achieve a higher performance. Thus, this ensemble of solutions will allow us to begin our analysis of this behavioral switch with a circuit whose dynamics can be fully visualized and then expand to circuits containing increasing numbers of interneurons.

Our second step will be to characterize separately the autonomous BBE dynamics of walking and chemotaxis with the best evolved 3-neuron circuit. In the case of chemotaxis, this should be straightforward because the entire BBE system is smooth and standard dynamical systems techniques can be applied [89, 128]. Although a number of software packages for such analysis exist, we have developed our own publicly-available *Mathematica* tools for dynamical systems analysis [15]. For walking, the nature of the legged body model means that the BBE system is only piecewise-smooth, requiring nonstandard techniques from hybrid dynamical systems [42]. However, we have extensive experience applying such techniques to the legged body model [14] and do not anticipate any insurmountable obstacles to this analysis. The end result of this step will be a characterization of the limit sets and their stabilities in the BBE chemotaxis and walking systems. The mechanisms underlying the limit sets corresponding to successful chemotaxis and walking behavior, and the dynamics of the transition between them when the body is switched, are what we wish to explain.

Our third step will be to characterize the autonomous dynamics of the highest-performing neural circuit parameterized by the sensory input S . Because CTRNNs are smooth dynamical systems, standard continuation techniques can be employed [89]. In addition, we have developed more specialized techniques for analyzing CTRNNs, including explicit expressions for some bifurcations, that can sometimes be employed [11; 13]. The end result of this step will be a bifurcation diagram of the S -parameterized CTRNN, giving the distinct phase portraits that are possible as S varies and the types of bifurcations that occur between them. If the sensory feedback that the circuit received from the body and environment in the full BBE system varied infinitely slowly, then this bifurcation diagram would provide a rigorous account of the circuit's nonautonomous dynamics. Unfortunately, since S typically varies on a timescale that significantly overlaps with at least some of the timescales of the circuit, such a quasistatic analysis provides only a rough guide to the role of the circuit dynamics in the full BBE systems generating walking and chemotaxis.

Our fourth step will be to characterize the nonautonomous dynamics of the best 3-neuron circuit in response to time-varying sensory input $S(t)$. Although mathematical frameworks exist for the analysis of nonautonomous dynamical systems [80, 111, 118], both the theoretical and practical state of nonautonomous dynamics is far less developed than for autonomous dynamics. We will proceed as follows. From the dynamics of the coupled BBE systems, we know the input signals that the circuit receives during chemotaxis and walking. We can abstract each of those signals into a parameterized family of closely related signals. For example, during walking the circuit receives periodic sensory input with a given period, amplitude, offset and shape. We can then embed this signal in a family of closely related signals using its first few Fourier modes and examine how the circuit's response changes as we move around within this family. A similar approach will be taken to the (usually nonperiodic) chemotaxis sensory signals. We

may also be able to define interpolations between these two families. Essentially, we are treating the circuit as a stateful nonlinear filter and probing its response to a given family of signals. Combining this analysis with the quasistatic analysis described in the previous step will give us a qualitative picture of the evolved circuit’s nonautonomous dynamics. As a more quantitative alternative, we will explore the use of methods such as dynamic mode decomposition (DDM), a dimensionality reduction technique that decomposes a time series into a set of oscillatory modes that are decaying or growing in time [120]. Since its original formulation, many applications and further developments have appeared [88], including its extension to input-output systems [113]. Techniques such as DDM will become increasingly important as the number of interneurons increases.

Our fifth step will be to use the above characterizations of the nonautonomous circuit dynamics to explain each behavior individually as well as the switch between them. For this, we will also need to characterize the nonautonomous dynamics of the two body models (how their dynamics transform their effector “inputs” from the neural circuit to their sensory “outputs” to the neural circuit). Given the relatively simple character of the body models, this is a much easier task than for the evolved CTRNNs. Then we will be able to explain how the two behaviors (as trajectories of the entire BBE system) arise from the reciprocal interaction between the nonautonomous dynamics of the circuit and that of the respective bodies. In particular, we will be able to understand in as much detail as we like the mechanistic underpinnings of different activity patterns that the highest performing 3-neuron circuit generates during the two behaviors, that is, its embodied multifunctionality.

Our sixth and final step will be to expand the insights gained from our detailed analysis of the best 3-neuron circuit to a larger set of evolved circuits. Much of the mechanistic detail underlying the operation of the best 3-neuron circuit will be specific to that circuit. The point of evolving an ensemble of circuits is precisely to have a larger set of high fitness models from which to generalize the results of our analysis. Thus, we will repeat the above analysis on the top nine remaining evolved 3-neuron circuits. We will look specifically for how similar performance is achieved in spite of whatever differences in nonautonomous dynamics, parametric autonomous dynamics, and neural and synaptic parameters are observed. Is the embodied multifunctionality of the top ten circuits grounded in the same principles at some level of abstraction, or does our explanation of multifunctionality in the best 3-neuron circuit need to be generalized? Then we will consider the top ten 4- and 5-neuron circuits. Can the same principles be applied when additional interneurons are present, or must our explanation be extended? What specific functional advantages do additional interneurons confer? Do larger circuits become more modular and specialized in their operation, or more distributed and multifunctional? The deliverable of this first specific aim will be a general understanding of how embodied multifunctionality plays out in this simple but representative example of body-triggered behavioral switching in a brain-body-environment system.

3.2 How can we identify and characterize the distinct information architectures that arise in embodied circuits?

A second fundamental challenge raised by embodied multifunctionality is describing the distinct functional circuits that arise from anatomically fixed neural circuits engaged in different tasks and dissecting their operation. To compute and analyze such emergent functional circuits, we will employ the tools of information theory. Specifically, we will focus on the mutual information that each neuron carries about an external stimulus and the pattern of information transfer between these neurons, using the visually-guided agent described earlier. We call such a characterization an *information architecture* (72). In general, information architecture is a time-varying feature of a brain-body-environment system engaged in a specific task. However, it can also be useful to average these quantities over time in order to obtain a static summary of the

key informational relationships. As we shall see, this specific aim differs from the first one along three dimensions: (1) the location of the change triggering the behavioral switch (environment vs. body), (2) the nature of the tasks being performed (visually-guided decision-making vs. motor control), and (3) the mathematical tools employed for analysis (information theory vs. dynamical systems theory).

The model setup involves two related tasks: object discrimination and perception of passability affordances. In the object discrimination task, a visually-guided agent must catch circular objects that are smaller than its own body, but avoid objects that are larger (Figure 4, left). In the passability task, the agent must align with apertures in a looming wall that are large enough for

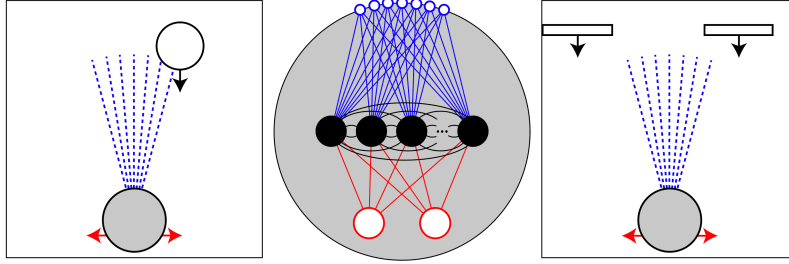


Figure 4: The same neural circuit (center) must discriminate objects size when it encounters circles (left) and passability when it encounters walls containing openings (right). Sensory inputs are shown in blue and motor outputs are shown in red.

its body to pass through, but avoid apertures that are too small (Figure 4, right). In both cases, we make use of the visually-guided body described earlier, which consists of an array of distal sensors projecting to a set of recurrently-connected interneurons, which in turn project to a pair of motor neurons driving the horizontal motion of the agent’s body (Figure 4, center). Note the duality of these two tasks.

Although they both involve decisions driven by relative size assessments, the tasks differ both in the nature of the size measured (positive space in the case of object discrimination vs. negative space in the case of passability) and the valence of a given size assessment (smaller sizes are “better” in object discrimination but larger sizes are “better” in the case of passability).

Our first step will be to evolve at least 100 agents that can solve both tasks. We will first evolve agents that can solve the tasks individually. We will identify the minimum number of interneurons necessary for an agent to perform each behavior successfully (99% fitness). The fitness evaluation for each of the tasks consists of 50 trials. During each trial, an agent receives a falling object (a circular object for the discrimination task and walls with an opening for the passability task) with a constant vertical velocity. Each trial differs in two respects: the size of the object/opening varies within a range; and the initial horizontal offset of the object/opening relative to the agent also varies within a range. We will maximize the average performance over all trials for each task, where performance is based on the following distances. In the case of object discrimination, the distance between the agent and the object is to be minimized for smaller objects and maximized for larger objects. For passability, the distance between the agent and the center of the opening is to be minimized for larger openings and maximized for smaller openings. When both tasks have been successfully evolved individually, we will evolve agents that can solve both tasks using the same neural circuit. Evolutionary runs with many different random seeds will be performed in order to systematically explore how agent performance depends on the number of interneurons. We will start with the minimum number of interneurons required to evolve agents for the individual tasks and then increase from there.

Our second step will be to identify the subset of neurons involved in each behavior in the best evolved circuit that results from the above set of evolutionary runs. We will compare two different methods to accomplish this goal: neural activity and mutual information. In the neural activity method, we will use the presence of activity in each neural trace during the different trials of the behavior as an indicator of involvement in the task [25]. To do this, we will record all neural and bodily state trajectories during all trials for each of the two tasks. Because the baseline of activity for the neurons in our model will have been artificially evolved, we will de-

fine the presence of neural activity during a task as the aggregated rate of change of a neuron’s activity over time. We will calculate the trial by trial rate of change of the neurons for each task. We will consider that a neuron belongs to the subcircuit for a certain task if it exhibits activity in *any* of the trials for that task. In contrast, the mutual information method involves characterizing the task-relevant information in each component of the circuit for each task. How much information does each neuron in the circuit have about the size of the circle during the object discrimination task and about the size of the opening during the passability task? To answer this question, we plan to calculate the mutual information $I(s; y)$ that each neuron y in the circuit has about the size s of the task-relevant information. Mutual information is a measure of the dependence between two random variables, X and Y ; it quantifies the amount by which a measurement on one of the variables reduces our uncertainty about the other [38]. We will estimate a joint probability distribution over values of each state variable and the task-relevant stimulus feature. We will consider the subcircuit for a task as the set of neurons with task-specific mutual information in any of the trials for that task. This analysis should reveal a finer subset of the neurons than observed in the initial neural-activity analysis. Having found the neurons involved in each behavior, we will examine how they are functionally connected.

Our third step will be to identify the static functional connectivity of the subcircuits involved in each behavior. We will compare three different methods to accomplish this goal: cross-correlation, transfer entropy, and information transfer. The importance of identifying task-specific subcircuits is that even if the subset of neurons involved in each task are the same, when the agent is engaged in different interactions with its environment the functional connectivity between that same subset of neurons could be different. We will start with the most frequently used measure to make inferences about the functional activity of a neural system: cross-correlational analysis of neural activity traces using Pearson’s correlation [1, 21, 53]. The idea is that the strength of the statistical relationship between the activity of every pair of neurons over time provides an index to the strength of their connectivity. We will consider the functional connectivity of a subcircuit for each task as the cross-correlation for every pair of neurons in the subcircuit averaged over trials. As a second method, we will use transfer entropy as the measure of functional connectivity [135]. Transfer entropy quantifies the information that one neuron X at time t transfers to a second neuron Y at time $t + 1$ by averaging over all time indices [121]. The advantage of using transfer entropy over correlational measures is that we can take into consideration non-linear relationships between neurons. We will consider the functional connectivity of a subcircuit for each task as the transfer entropy for every pair of neurons in the subcircuit averaged over trials. The third method will allow us to identify the unique pathways along which task-specific information transfers within the subcircuits. We can use multivariate extensions of information theory to determine the information transferred between two neurons about task-relevant stimuli [141]. We will measure information transfer as the amount of information that is transferred from one neuron Y to another Z about the task relevant stimuli, X , calculated as the information about X shared by Z_t and Y_{t-1} that was not already in Z at time $t - 1$. Information transfer requires the use of a measure of redundant information between two variables. We will consider the different methods for calculating redundant information that have been proposed in the literature [76]. We will consider the functional subcircuit for a task as the time-averaged and trial-averaged information transfer between each pair of neurons for that task. We expect that each of the methods should reveal an increasingly refined task-specific difference in the functional connectivity of the subcircuits.

Steps two and three provide us with the static information architectures for each task. In order to understand how multifunctionality arises in these brain-body-environment systems, we have to consider how the information architecture changes over time with behavior. Our fourth step will be to characterize the dynamics of the task-specific information architectures that arise as the agent interacts with the different environments. We will characterize the time-varying information architecture for each task by unrolling mutual information for each of the neurons

in the circuit across time, and by unrolling the information transferred between pairs of neurons also over time. In order to do this, we will estimate joint probability distributions over values of each state variable and task-relevant stimulus feature at each moment in time. The time-varying distributions of random variables will be estimated over a fine grid of possible values and times using a kernel density estimation technique known as average shifted histograms [122]. We will determine a sufficiently high number of trial presentations, a fine enough grid of bins, and a reasonable number of shifts for the histograms along each dimension, and we will verify that the results of our analysis are qualitatively robust over a wide range of trials, resolutions, and shifts. The focus of the analysis will be to better understand how different components of the neural circuit turn into either specialized components for a specific task or into components that are reused over the multiple tasks, based on the BBE interaction. Furthermore, a dynamic information architecture should allow us to examine whether components of a neural circuit change their degree of specialization or reuse at different times during the task.

Our fifth step will be to dissect the dynamic information architectures based on the specific information that each neuron in the circuit has about the relevant stimuli. While a circuit may show similar subsets of neurons involved in different tasks, similar functional connectivity, and even similar dynamics in the information architectures, ultimately what specific information a neuron carries about the task-relevant stimuli is what determines the particular role that it plays in guiding behavior. Mutual information averages across all measurement outcomes. We can use specific information to unroll such averages, quantifying the information that each neuron has about each possible outcome of a stimulus measurement (in this case, the size of the object/opening) [28, 41]. We will focus our analysis specifically on neurons that exhibit a substantial amount of reuse across the different tasks. In neurons that share information about the size of the opening in one task and about the size of the object in the other task, what is the specific information that they have about each? Do some neurons specialize on detecting smaller sizes or larger sizes in both tasks? Furthermore, we can unroll specific information over time as well in order to understand how the specific role of the neurons change dynamically as the entire BBE engages in different tasks [18].

Our sixth step will be to characterize how task-relevant information is distributed across the different components of the subcircuits for each task. When analyzing information in a system, it is necessary to consider not only the information carried by individual variables about the task-relevant stimuli, but also the information that may be encoded uniquely, redundantly, and synergistically by multiple variables. When we consider multiple neurons simultaneously (e.g., pairwise or higher), how do the different neurons encode information about the task-relevant issues? To answer this question, we will use the partial information decomposition method described above [140], where unique contributions correspond to the information provided by one source that is not available in the other source; redundant contributions correspond to information available in both sources; and synergistic contributions correspond to the information provided by the simultaneous knowledge of both sources that is not available from either source alone. For each pair of neurons in the subcircuit, we will ask: do the neurons encode information about the task-relevant stimulus uniquely, redundantly, or synergistically?

Our seventh and final step will be to expand the insights gained from our analysis of the best multifunctional visually-guided agent to a larger set of evolved circuits. To this point, we have examined in detail the pattern of information architecture of the best-performing multifunctional circuit only. The complete ensemble of multifunctional circuits is likely to contain many others with comparable performance. We will focus on the highest-performing subset of the ensemble, namely those networks having a combined performance of over 95% on both tasks. We will use what we learned from the individual analysis to identify the task-specific information architectures for the ensemble of high-performance solutions. We will examine the degree of task-specific neuron reuse overlap across the diverse set of successful multifunctional circuits. We will partition the set of evolved circuits according to whether the subcircuits are

non-overlapping, different degrees of partially-overlapping circuits, and fully-overlapping. We will also use the static information architecture of each solution, encoded as a feature vector, to identify clusters in the ensemble. Finally, once we have a good understanding of the ensemble of the multifunctional circuits, we will compare it with the ensemble of circuits that were trained to solve only each of the individual tasks. The main question of interest we will examine is whether circuits that perform multiple behaviors encode information differently than circuits that perform single behaviors. The deliverable of this second specific aim will be an information theoretic understanding of how embodied multifunctionality plays out in visually-guided agents undergoing an environment-triggered behavioral switch.

3.3 *How does embodied multifunctionality scale with behavioral complexity?*

Our final aim is to examine how the theoretical insights into embodied multifunctionality obtained from our first two specific aims scales with behavioral complexity. A number of factors can contribute to increased behavioral complexity, including larger neural circuits, bodies with higher degrees of freedom, bodies with more sensors and different sensory modalities, richer environments, and more demanding tasks. Such factors, both singly and in combination, necessitate a richer behavioral repertoire containing a larger number of different behaviors with more complicated conditions for switching. Most previous research in evolutionary robotics has considered only one or two behaviors at a time. In order to move beyond this state of the art, we must address two crucial methodological challenges: (1) scaling our ability to evolve successful brain-body-environment systems that can fluidly shift between different behaviors and (2) scaling our ability to analyze the operation of such systems. We will engage these challenges by incrementally evolving and analyzing a sequence of increasingly more complicated BBE systems and testing the extent to which they operate according to the same principles revealed in specific aims 1 and 2.

The first step of our final aim is to design a BBE model with the flexibility to easily scale up in complexity. We will build on the visually guided agent studied in specific aim 2. By increasing the complexity of the environment, the complexity of the tasks the agent can face will begin to grow. For example, if the objects from the task proposed in specific aim 2 have color intensity to them, then it is possible to create an additional context-dependence to the task/behaviors. If the objects are black, then the task is the same as in specific aim 2; if the objects are white, then the tasks are inverted with respect to whether to approach or avoid as a function of the relative sizes. A second example involves an environment where as soon as the agent begins to move, the objects disappear from the agent’s field of view. Such an extension challenges the agent to transcend its immediate environment by allowing past experiences to influence its future actions. In all of these conditions, the specific nature of the task the agent has to solve has to be inferred through the agent’s interaction with the environment. With each new extension to the BBE system, the number of behaviors the agent has to be capable of switching to increases combinatorially. This combinatorial set up of the behaviors allows the model to grow in complexity in a tunable way (for a similar approach in a disembodied agent, see [144]). Accordingly, the sensor and motor capacities of the agent will be increased as necessary to solve the tasks in those environments. In the example given, this would require the addition of a basic form of color vision. More generally, this can include extending the model to have higher-resolution vision, multiple eyes, the ability to foveate, the addition of new sensory modalities, the addition of arms and legs, the ability to move in two dimensions, proprioception, etc. As the complete BBE system grows in complexity, the agent should be able to accomplish tasks over a broad range, including many of the areas that have previously been explored only individually (e.g., object manipulation, locomotion, navigation, learning and memory, and communication).

Our second step is to systematically compare different strategies for evolving agents to perform multiple tasks. Our previous and preliminary work suggests two tasks can be evolved us-

ing standard evolutionary techniques. As the number and complexity of the behaviors increase, there are several complementary strategies that have the potential to facilitate success: incremental task shaping, incremental complexity shaping, incremental neural complexity, modular neural architectures, and multi-objective optimization. First, we can gradually increase the number of behaviors agents have to switch between over evolutionary time, evolving agents to solve one of the behaviors first, and then incrementally add new ones when agents succeed at the current ones. This incremental shaping protocol has been successful in the past [70, 110]. Second, we can start with simpler versions of all the behaviors and then gradually increase their complexity in parallel. Third, we can gradually increase the amount of neural resources available to the controller over evolutionary time. Whenever the population becomes stuck without making progress for a certain period of time, we can add neurons to the circuit, opening up new directions in parameter space for the evolutionary algorithm to explore. Similar approaches have been described in related literature [32, 126, 148]. These three strategies involve establishing a criterion (typically a performance threshold), such that when it is met by the population of evolving agents, the BBE system automatically changes. Fourth, we will examine different neural architectures that facilitate the evolution of multifunctional circuits. Typically, recurrent neural networks used in embodied tasks are fully connected. We will examine the benefit of modular neural architectures by evolving circuits with different degrees of community structure [100]. Finally, we will examine the use of multi-objective optimization algorithms [40]. In most work on evolving dynamical recurrent neural networks to solve multiple embodied tasks [31, 69], the performance of the agent on each task is combined into a single fitness value, either additively or multiplicatively (or sometimes using the minimum of the different performance scores). With multi-objective optimization methods, the different performances are not combined; instead individuals with high performance across the complete Pareto front are maintained. In preliminary experiments, we have successfully implemented and tested evolving BBE systems using multi-objective optimization strategies. An additional advantage of using a multi-objective optimization technique is that we can focus our analysis on an ensemble of solutions that are spread across the full spectrum of the Pareto optimum of the multiple tasks.

Our third step is to determine the effective dimensionality of the neural circuits as the number of behaviors increases. Does multifunctionality require summing the effective dimensionality of each task independently, or are there synergistic effects, such that, as the number of behaviors goes up, the effective dimensionality stays relatively constant? We will compare the effective dimensionality of circuits that have been trained to perform a gradually increasing number of tasks. Effective dimensionality was first introduced in physics [19] and subsequently employed in neuroscience [117]. To quantify the dimension of the subspace containing the neural trajectory involved in a certain set of behaviors, we will calculate $d = 1/\sum_i \tilde{\lambda}_i^2$, where $\tilde{\lambda}_i = \lambda_i/\sum_j \lambda_j$ are the normalized eigenvalues of the covariance matrix of the points in the neural trajectory. The quantity can be roughly thought of as the number of principal components needed to capture most of the variance of the points, and is maximized when all eigenvalues are equal, and minimized when one eigenvalue is much larger than the rest. In addition to determining synergistic effects, effective dimensionality could be a way to succinctly assess the degree of specialization or reuse in the different subcircuits responsible for each of the different behaviors. The lower the effective dimensionality of a multifunctional circuit, the more likely its neural resources are being reused across behaviors. We can further validate this idea by identifying and characterizing the information architectures of circuits with different effective dimensionalities. Effective dimensionality will also give us some indication of the difficulty to be expected in attempting to analyze a given multifunctional BBE system.

Our fourth step will be to examine the operation of these richer examples of embodied multifunctionality. Given their complementary strengths and weaknesses [18], we will tackle this challenge by combining the dynamical systems and information theoretic approaches from aims 1 and 2, respectively. Specifically, we will first employ information theory to identify and char-

acterize the information architecture operating during each behavior and track the dynamics of its information flow throughout the BBE system. This analysis will provide us with a set of guideposts, highlighting the key subsets of neurons and connections and the informational roles that they play in each behavior. How does size vs. color information flow through the BBE system? Where is the “memory” of a previously-experienced but no longer visible object stored? We will then turn to a dynamical systems analysis to answer such focused questions, as described in specific aim 1. We believe that combining the descriptive power of information theory with the causal insight provided by dynamical systems theory can be a powerful tool for revealing the mechanisms by which embodied multifunctionality operates.

Our final step is to test how the theoretical insights into embodied multifunctionality gained from our studies in specific aims 1 and 2 generalize to the more complex cases of multiple behaviors that we will evolve. As we study the functioning of these more complicated BBE systems, do we see the same principles in operation? Or do new possibilities arise? If the latter, what new features do we observe? More generally: Can larger-scale multifunctional systems be understood as a subset of nonautonomous dynamics that are evoked from neural circuits when they are embedded in a given sensorimotor feedback loop through its body and environment? Can the task-specific subcircuits that are involved in each of the distinct behaviors be identified from an analysis of the neural traces during behavior? What are the conditions that lead to multifunctional systems with specialized subcircuits versus ones with fully- or partially-overlapping subcircuits? For example, an argument could be made that larger nervous systems in relation to the number of behaviors required might provide more opportunities for synergistic interactions between neurons, and therefore more overlap; however relatively larger nervous systems might also provide more space for neural circuits to become independently specialized. In any case, our attempt to scale the behavioral complexity of our approach should reveal important insights into how animals – and potentially artificial systems – can achieve the rich, flexible behavioral repertoires that they exhibit with fixed neural resources.

II. Broader Impacts

The main goal of our proposed work is to extend the notion of multifunctionality (aka polymorphism, reorganization, reuse) to the fully situated and embodied context of an entire brain-body-environment system. We view the primary contribution of this line of research as theoretical: How are we to understand the role of multifunctional neural activity within anatomically fixed circuits in the generation, coordination and switching of multiple behaviors? We expect the results of such an effort to be of interest not only within computational and theoretical neuroscience, but within neuroscience more broadly [4, 7-8, 29-30, 56, 83, 86, 99, 144-145]. In addition, insights obtained from the proposed research should be directly applicable to the study of habits within the sensorimotor contingency approach of enactive cognitive science [27; 47]. The results should also be of interest in technological applications, including work within AI utilizing evolutionary search [127; 129], work within machine learning on embodied control tasks and continual learning [63; 107], work on multiple behaviors in evolutionary robotics [48; 97; 124], and even work on multiple behaviors in disembodied agents [144]. In order to facilitate these and other applications, we intend to make the code for both models and analysis tools publicly available upon publication of the associated papers.

The PI and co-PI’s labs have a long history of interdisciplinary training and outreach. The proposed work will provide opportunities for interdisciplinary training for two graduate students. We work with undergraduate students through the Indiana University Cox Research Scholarship Program, which supports talented Indiana residents with demonstrated financial need to participate in a research project, the Center for Excellence for Women in Technology, and the Undergraduate Research Opportunities in Computing program. The PIs also teach both undergraduate and graduate courses that incorporate techniques and examples from their re-

search. Perhaps the most notable examples include a popular undergraduate laboratory course on Autonomous Robotics from the perspective of embodied cognitive science taught by the PI and a freshman course on Minds and Machines taught by the co-PI that is one of a series aimed at introducing freshmen to a range of possible STEM careers. Finally, both PIs regularly discuss their research with such local undergraduate groups as the Student Organization for Cognitive Science, the Undergraduate Neuroscience Student Group, the Women Who Code Special Interest Group, and the Center for the Integrative Study of Animal Behavior’s Summer REU Program in Animal Behavior. In terms of graduate students, both PIs are core members of an NSF/NRT in Complex Networks and Systems at IU (PI: Luis Rocha) and they contribute to NRT training activities and provide lab rotations and research homes for NRT students. With respect to outreach and public understanding of science, the PIs participate in the Bloomington Science Café, the Indiana University Science Fest, and the Indiana University Robotics and Intelligent Systems Open House. Our labs also directly interact with members of the public about science, including local retirement communities, scouting troops and schools. All such training and outreach activities are closely linked with our currently funded research projects, and that will certainly continue to be the case with the work proposed here.

III. Results from Prior NSF Support

“An Ensemble of Neuromechanical Models of *C. elegans* Locomotion,” R.D. Beer (PI) and E.J. Izquierdo (co-PI), IIS-1524647, \$492,189 total costs, 8/1/2015-8/31/2019. **Intellectual Merit:** The overall goal of this grant was to construct and analyze biologically-constrained neuromechanical models of forward locomotion in the nematode worm *C. elegans*. Specifically, we (1) developed a new optimized implementation of the Boyle-Berry-Cohen (BBC) [24] mechanical model of the nematode body with updated musculature, stretch receptors and neural model that was suitable for optimization with evolutionary algorithms, (2) showed that an intrinsic network oscillator consistent with a statistical summary of ventral cord connectivity in the worm [62] could generate locomotion-like dorsal-ventral oscillations, (3) demonstrated that a chain of such circuits coupled via identified gap junctions from the *C. elegans* connectome and embodied in the BBC body model can produce forward locomotion whose characteristics are consistent with those of the worm even in the absence of stretch receptor feedback, (4) analyzed the mechanisms by which the oscillations were generated and propagated, and (5) also developed a second model driven by a network oscillator in the head and propagated by stretch receptors along the body. Research funded by this grant produced a total of 7 journal and peer-reviewed conference publications [31, 65-67, 105-106, 146]. **Broader Impact:** Interdisciplinary training was provided to a postdoctoral associate, including mentoring in teaching, presenting, and paper and grant writing. Opportunities were also provided for local, national and international presentation. In addition, our labs have also been active in outreach activities, including incorporating techniques and examples from ours results in undergraduate teaching and public presentations. We also began training a new graduate student in this area, as well as an undergraduate through the IU Cox Research Scholars Program. The work has been presented at a number of workshops and conferences, including the Computational Neuroscience Meeting, Computational and Systems Neuroscience, Neural Information Processing Systems, Society for Neuroscience, the International *C. elegans* Conference, the SIAM Conference on Applications of Dynamical Systems, and the Cognitive Science Conference. Finally, we have made our *C. elegans* neuromechanical model available on GitHub for use by other researchers [68].

In addition, co-PI Izquierdo just received a CAREER award (1845322) to study the neural coordination between two distinct spatial-navigation strategies that run in parallel in the nematode worm *C. elegans* through the use of biologically-grounded neuromechanical models. Any methods and insights developed in the current proposal for understanding embodied multifunctionality will be immediately applicable to the *C. elegans* work.