Check for updates

Empowerment As Replacement for the Three Laws of Robotics

Christoph Salge[1,2] and Daniel Polani[1]*

[1] *SEPIA Unit, Adaptive Systems Research Group, School of Computer Science, University of Hertfordshire, Hatfield, United Kingdom,* [2] *Game Innovation Lab, Department of Computer Science and Engineering, New York University, Brooklyn, NY, United States*

The greater ubiquity of robots creates a need for generic guidelines for robot behavior. We focus less on *how* a robot can technically achieve a predefined goal and more on *what* a robot should do in the first place. Particularly, we are interested in the question how a heuristic should look like, which motivates the robot's behavior in interaction with human agents. We make a concrete, operational proposal as to how the information-theoretic concept of *empowerment* can be used as a generic heuristic to quantify concepts, such as self-preservation, protection of the human partner, and responding to human actions. While elsewhere we studied involved single-agent scenarios in detail, here, we present proof-of-principle scenarios demonstrating how empowerment interpreted in light of these perspectives allows one to specify core concepts with a similar aim as Asimov's Three Laws of Robotics in an operational way. Importantly, this route does not depend on having to establish an explicit verbalized understanding of human language and conventions in the robots. Also, it incorporates the ability to take into account a rich variety of different situations and types of robotic embodiment.

Keywords: empowerment, information-theory, robot, intrinsic motivation, human robot interaction, value function, three laws of robotics

1. INTRODUCTION

One of the trends of modern robotics is to extend the role of robots beyond being a specifically designed machine with a clearly defined functionality that operates according to a confined specification or safely separated from humans. Instead, robots increasingly share living and work spaces with humans and act as servants, companions, and co-workers. In the future, these robots will have to deal with increasingly complex and novel situations. Thus, their operation will require to be guided by some form of generic, higher instruction level to be able to deal with previously unknown and unplanned-for situations in an effective way.

Once robotic control has to cope with more than replaying meticulously pre-arranged action sequences, or the execution of a predefined set of tasks as a reaction to a specified situation, the need arises for generic yet formalized guidelines which the robot can use to generate actions and preferences based on the current situation and the robot's concrete embodiment.

We propose that any such guidelines should address the following three issues. First, "robot initiative": we expect the principles to be generic enough for the robot to be able to apply them to novel situations. In particular, the robot should not only be able to respond according to predefined situations but also be able to *generate* new goals and directives as needed in new situations. Second, breaking action equivalence: how should a robot choose between several different actions when all

produce essentially the same desired outcome, only in different ways? Can we formulate good secondary criteria that the robot should optimize in addition, once it can ensure that the primary job gets done? Finally, safety around robots: the default approach often involves a "kill switch," i.e., the drastic and crude step of shutting down the robot and stopping all its actuators. This rudimentary response is often undesirable, e.g., when the robot is carrying out a vital function where an immediate shutdown would lead to harm a human or itself, or when, to maintain safety and prevent damage, the robot is required to *act* rather than to *stop acting*. In summary, we propose that there is a pronounced need for generic, situation-aware guidelines that can inform and generate robot behavior.

Science Fiction literature, an often productive vehicle to explore ideas about the future, has come across this problem in its countless speculations about the role of robots in society. Arguably, the best-known suggestion for generic rules for robot behavior are Asimov's *Three Laws of Robotics* (Asimov, 1942):

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given to it by human beings, except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

While there is ample room to discuss the technicalities and implications of the Three Laws (McCauley, 2007; Anderson, 2008; Murphy and Woods, 2009), we believe most people would agree with the general sentiment of the rules; Asimov himself argued that these rules are not particularly novel, but govern the design of any kind of tool humanity produces (Asimov, 1981). Asimov stated that he aimed to capture basic requirements of tools, namely, safety, compliance, and robustness, and his Three Laws are an attempt to explicitly express these properties in language. But one central problem in adapting these rules to current-day robots is the scant semantic comprehension of rules expressed in natural language.

In part, this is based on fundamental AI problems, such as determining the scope and context pertinent to such rules by robots or AIs in general (Dennett, 1984). Another AI-philosophical problem raises the question on how to assign meaning to the semantic concepts (Coradeschi et al., 2013). Since robots usually have a radically different perspective to humans, and hence a different perceptual reality, it remains doubtful if robots and human could have a common language. Already simple, and common concepts, such as "harm," cannot be naively related to the robot's perspective. Subsequently, it is difficult to build robots that understand what constitutes harm and thus can avoid inflicting it.

Even if we were to somehow imbue a robot with a human-level understanding of human language, we would still face the more pragmatic problem that human language carries intrinsic ambiguity. One example of this is the legal domain, where humans will argue what exactly constitutes "harm" in legal cases, demonstrating that there is no unambiguous understanding of this term that could just be applied in a technical fashion. Relatively simple sentences, such as the amendments of the US constitution have

spawned decades of interpretation. Also, several of Asimov's stories illustrate how robots find loopholes in their interpretations of the Three Laws that defy human expectations. Because of all previously mentioned problems, current-day robots are unable to generate actions or behavior complying with natural language directives, such as "protect human life" or "do no harm." Even greater demand in regard to natural language processing is posed by the Second Law which requires the robot to be able to interpret any order. This requires a robust, unambiguous understanding of human language that cannot even be realized by humans.

In this paper, we have a similar aim as Asimov had with his Three Laws; however, rather than exactly reproduce the Laws, we propose a formal, non language-based method to capture the underlying properties of robots as tools. Instead of employing language, we suggest to use the information-theoretic measure of empowerment (Klyubin et al., 2008) in particular, and potential causal information flow (Ay and Polani, 2008) in general, as a heuristics to produce characteristic behavioral phenomenologies which can be interpreted as corresponding to the Three Laws in certain, crucial aspects. Note that we do not expect to reproduce the precise behavior associated with the Three Laws, but rather to capture essential intuitions we have about how the Three Laws should operate; those which cause us to agree with them in the first place. Importantly, we do not argue that the proposed heuristics give a complete and sufficient account for ethical robot behavior. Rather, at this stage, we consider properties such as safety, compliance, and robustness as secondary to the main robot mission. This is very much in line with the idea of robots as servants and companions put forward in the "principles of robotics" (Boden et al., 2011). In contrast, principles such as Tilden's Laws of Robotics that focus solely on the robot as autonomous, self-preserving life forms—basically living machines (Hasslacher and Tilden, 1995)—are expressly not in the scope of this article. Nevertheless, we like to point out existing work that links empowerment to the idea of autonomous and living systems (Guckelsberger and Salge, 2016).

Centrally, we propose here that the empowerment formalism offers an operational and quantifiable route to technically realize some of the ideas behind the Three Laws in a generic fashion. To this end, we will first introduce the idea behind empowerment. We will then proceed to give both a formal definition and the different empowerment perspectives. We will then discuss how these different perspectives correspond to concepts, such as self-preservation, compliance, and safety. Finally, we will discuss extensions, challenges, and future work needed to fully realize this approach on actual robots.

2. EMPOWERMENT

Empowerment is an information-theoretic quantity that captures how much an agent is in control of the world it can perceive. It is formalized by the information-theoretic channel capacity between the agent's actuations during a given time interval and the effects on its sensory perception at a time following this interval. Empowerment was introduced by Klyubin et al. (2005) to provide agents with a generic, a prioristic intrinsic motivation that might act as a stepping stone toward more complex behavior.

An information-theoretic measure, it quantifies how much *potential* causal influence an agent has on the world it can perceive. Empowerment was motivated by the idea to unify several seemingly disparate drives of organisms of multiple levels of complexity, such as maintaining a good internal sugar level, staying healthy, becoming a leader in a gang, accumulating money, etc. (Klyubin et al., 2008). While all these drives enhance survivability in one way or another, one unifying theme that ties them together is maintaining and enhancing one's ability to act and control the environment. Empowerment attempts to capture this notion in an operational formalism; in this paper, we specifically want to demonstrate how this principle can serve as a cognitive heuristic to generate behavior in the spirit of the Three Laws of Robotics. For this, we consider empowerment from different, but related perspectives.

To motivate empowerment and gain a better understanding, let us first take a brief look at the background, before moving on to the formal definition. Oesterreich (1979) argues that agents should act so that their actions lead to perceivably different outcomes, which he calls "efficiency divergence" ("Effizienzdivergenz" in German). In the ideal case, different actions should lead to different perceivable outcomes. Von Foerster (2003) famously stated "I shall act always so as to increase the total number of choices," arguing that a state where many options are open to an agent is preferable. Furthermore, Seligman (1975) argues that humans who are forced to be in a state where one's actions appear to have random outcomes or no outcome variation at all suffer mental health problems. This relates to more recent empirical studies by Trendafilov and Murray-Smith (2013), which indicated that humans in a control task associate a low level of empowerment with frustration and perform better in situations where they are highly empowered. More recently, ideas similar to empowerment have also emerged in physics (Wissner-Gross and Freer, 2013), proposing a closely related action principle; the latter is, however, motivated by the hypothesized thermodynamic Maximum Entropy Production Principle instead of being based on evolutionary and psychological arguments.

2.1. Empowerment As Intrinsic Motivation

In essence, empowerment formalizes a "motivation for effectance, personal causation, competence, and self-determination," which is considered to be one area of intrinsic motivation by Oudeyer and Kaplan (2007), Oudeyer et al. (2007). Intrinsic motivation is a term introduced by Ryan and Deci (2000) as: "[…] the doing of an activity for its inherent satisfaction rather than for some separable consequence." In the last decades, a number of methods have been suggested to artificially generate behaviors of a similar nature as has been hypothesized about intrinsically motivated organisms. Among these are *Artificial Curiosity* (Schmidhuber, 1991), *Learning Progress* (Oudeyer and Kaplan, 2007), Predictive Information (Ay et al., 2008), *Homeokinesis* (Der et al., 1999), or the *Autotelic Principle* (Steels, 2004). Not all of them are related to personal causation, most are more focused on the "reduction of cognitive dissonance" and on "optimal incongruity." But all of them are to some extent intrinsic to the idea of agency itself, and they all share a set of properties that makes them well suited to imbue agents and robots with generic, motivated behavior

without reliance on an externally defined reward structure. They usually are

- task-independent,
- computable from the agent's perspective,
- directly applicable to many different sensorimotor configurations, without or with little external tuning, and
- sensitive to and reflective of different agent embodiments.

The task-independence demarcates this approach from most classical AI techniques, such as reinforcement learning (Sutton and Barto, 1998); the general idea is not to solve any *particular* task well, or to be able to *learn* how to do a specific task well, but instead to offer an incentive for behavior even if there is currently no specific task the agent needs to attend to. In the case of empowerment, the behavior generated turns out to coincide well with the idea of robot self-preservation.

The computability from an agent's perspective is an essential requirement. If some form of intrinsic motivation is to be realized by an organism or deployed onto an autonomous robot, then the organism/robot needs to be able to evaluate this measure from its own perspective, i.e., based on its own sensor input. This relies on a notion of *Umwelt* by von Uexküll (1909), because any intrinsic preference relation will be defined with respect to the agent's experience, i.e., its perceived dynamics between the agent's sensors and actuators; the latter, in turn, arises from the interplay of the environment and the embodied agent. This is closely related to the concept of a "counterworld" (Gegenwelt (von Uexküll, 1909)), the internal mirroring of the Umwelt, but one that only captures functional circles, those relations where actions relate to relevant feedback from the environment. Empowerment fits with this approach, as it does not require an explicit, full world model that tries to capture the whole environment, but only a short-term forward model that relates its current actions and context (typically a sensor-based belief about the current state) to the subsequent sensor states which are expected to result from these actions. Ziemke and Sharkey (2001) provide a more in-depth explanation of the Umwelt concept and also provide an overview of modern robotic approaches, such as the Subsumption Architecture by Brooks (1986) that realizes robot control as a hierarchy of functional circles. Empowerment, and to a large extent also the other mentioned intrinsic motivation measures, are usually compatible with these bottom-up approaches with minimal models focused on immediate action–perception loops.

The next property of intrinsic motivation is the ability to cope with different, quite disparate sensorimotor configurations. This is highly desirable for the definition of general behavioral guidelines for robots. This means not having to define them separately for every robot or change them manually every time the robot's morphology changes. The applicability to different sensorimotor configurations combined with the requirement of task-independence is the central requirements for such a principle to be universal. More precisely: to be universal, a driver for intrinsic motivation should ideally operate in essentially the same manner and arise from the same principles, regardless of the particular embodiment or particular situation. A measure of this kind can then identify "desirable" changes in both situation (e.g., in the context of behavior generation) and embodiment

(e.g., in the context of development or evolution). For example, while most empowerment work focusses on state evaluation and action generation, some work also considers its use for sensor or actuator evolution (Klyubin et al., 2008).

Furthermore, this implies that a measure for intrinsic motivation should not just remain generically computable, but also be sensitive to different morphologies. The challenge is to define a value function in such a way that it stays meaningful when the situation or the embodiment of the agent changes. An illustrative example here are studies where an agent had the ability to move and place blocks in a simulated world (Salge et al., 2014a). Driven by empowerment maximization, the agent changed the world in such a way that the resulting structures ended up reflecting the particular embodiment of the agent.

To sum up this section, if we want to use an intrinsic motivation measure as a surrogate for what Pfeifer and Bongard (2006) call a value function in the context of embodied robotics, then we propose it needs to fulfill these criteria. Here, we specifically concentrate on empowerment, since it has been shown to be a suitable candidate to produce the desired behavior; furthermore, many of its relevant properties have already been studied in some detail now (Salge et al., 2014c). However, we would like to emphasize that the concepts to be developed below are not limited to empowerment in their application, but might also be adaptable to alternative intrinsic motivation methodologies in order to provide operational safety principles for robots.

2.2. Empowerment Formalism

To make our subsequent studies precise, we now give a formal definition of empowerment. Empowerment is formalized as the maximal potential causal flow (Ay and Polani, 2008) from an agent's actuators to an agent's sensors at a later point in time. This can be formalized in the terms of information theory as channel capacity (Shannon, 1948).

To compute empowerment, we model the agent–world interaction as a perception–action loop as in **Figure 1**. In **Figure 1**, we are looking at a time-discrete model, where an agent interacts with the world. An agent chooses an action A_t for the next time step based on its sensor input S_t in the current time step t. This influences the state of R_{t+1} (in the next time step), which in turn influences the sensor input S_{t+1} of the agent at that time step. The cycle then repeats itself, with the agent choosing another action in A_{t+1}. Note that, in a more general model, this choice of action

might also be influenced by some internal state of the agent which carries information about the agent's past.

For the computation of empowerment, we consider this perception–action loop as telling us how actions may potentially influence a state in the future, and by influence we emphatically mean not the actual outcome of the concrete trajectory that the agent takes, but rather the potential future outcomes at the given time horizon $t + 3$, i.e., the distribution of outcomes that could be generated by actuation, starting from time t. The most straightforward interpretation is as a probabilistic communication channel where the agent transmits information about actions A_t, A_{t+1}, A_{t+2} through a channel and considers how much of it is reflected in the outcome in S_{t+3}. The maximal influence of an agent's actions on its future sensor states (again, not its actual actions, but its *potential* actions) can now be modeled formally as Shannon channel capacity: what the agent may possibly (but need not) transmit over the channel—or, in fact, what the agent may (but need not) have changed in the environment, at the end of its 3-step action sequence.

Empowerment is then defined as the channel capacity between the agent's actuators A in a sequence of time steps and its own sensors S at a later point in time. For example, if we look at empowerment in regard to just the next time step, then empowerment can be expressed as

$$\mathfrak{E}(r) := C(A_t \to S_{t+1} \mid r) \equiv \max_{p(a_t|r)} I(S_{t+1}; A_t \mid r). \quad (1)$$

Note that the maximization implies that it is calculated under the assumption that the controller which chooses the action sequences $(A_t)_t$ is completely free to act from time t onward (but permitted to correlate the actions) and is not bound to a particular behavior strategy $p(a|s, r)$.

Furthermore, empowerment is a state-dependent quantity, as it depends on the state r in which the consideration of potential futures is initiated. Instead of an externally observable objective state r, one could also consider empowerment based on a purely agent-intrinsic "context" states derived from earlier agent experiences (Klyubin et al., 2008) without conceptual changes. For simplicity and clarity, we will here discuss only empowerment landscapes depending on "objective" states r. We also assume for the beginning of the paper that the agent has somehow previously acquired a sufficiently accurate local forward model $p(s_{t+1}|a_t, s_t)$.
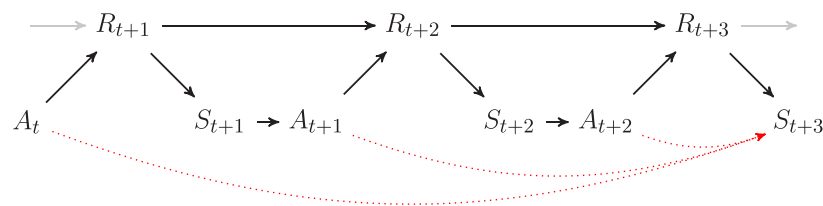

FIGURE 1 | The perception–action loop visualized as a Causal Bayesian network (Pearl, 2000). S is the sensor, A is the actuator, and R represents the rest of the system. The index t indicates the time at which the variable is considered. This model is a minimal model for a simple memoryless agent. The red arrows indicate the direction of the potential causal flow relevant for 3-step empowerment. If empowerment is measured, the input to each of the actions is freely chosen, disconnected from the sensor input, but inside the 3-step horizon the actions may expressly be correlated—see also discussion in Klyubin et al. (2008).

We will later discuss the general difficulties and implications of the model acquisition itself.

Empowerment is defined for both discrete and continuous variables. However, while it is possible to directly determine the channel capacity for the discrete case using the Blahut–Arimoto Algorithm (Arimoto, 1972; Blahut, 1972) to compute a channel capacity achieving distribution $p^\star(a|r)$, this algorithm cannot be straightforwardly applied to the continuous case. There exist adaptations for empowerment calculation in the continuum, though. Jung et al. (2011) use Monte-Carlo Integration to approximate empowerment, but this method is very computationally expensive. A significantly faster method approximates empowerment of a continuous channel by treating it as a linear channel with added independently and identically distributed (i.i.d.) Gaussian noise (Salge et al., 2012).

2.3. Empowerment Maximization Behavior

When talking about an empowerment-maximizing agent, it must be emphasized that the distribution $p^\star(a|r)$ *achieving* the channel capacity is *not* the one that an empowerment-maximizing agent *actually* uses to choose its next action. The capacity-achieving distribution is only utilized to compute the empowerment value of world states; for each such state, an empowerment value is thus computed, which defines a pseudo-utility landscape on the state space. For the actual action, the agent chooses a greedy strategy with respect to this pseudo-utility: it chooses its actions as to locally maximize the empowerment of the next state it will visit.

To illustrate, an agent might prefer a state where it would have the *option* to jump off a cliff. However, assuming that the agent would break from the fall or have to tortuously climb up again, the agent would not actually select the action that would cause it to fall off the cliff; it will just want to possess the option. Importantly, this requires the agent to have a correct forward model of what will happen when stepping off the cliff. If there is noise in the dynamics, this will, on the other hand, lead empowerment to pull the agent slightly away from the cliff, as the agent in this case

cannot ensure that it would not accidentally fall off the cliff. This is similar to on-policy reinforcement learning.

In past work, two main strategies have been used. Greedy empowerment maximization basically considers all possible actions in the current state and then computes the successor states (or state distributions) for each of those actions. Then, empowerment is calculated for each of those successor states. The successor state with the highest empowerment is selected, and the agent then performs the action leading to the chosen successor state. In case each action has a distribution of successor states, then the action with the highest average successor state empowerment is selected. This has the advantage that the agent only needs to compute the empowerment for the immediate successor states in the future.

Alternatively, the agent could compute the empowerment values for each possible state of the world, or a subset thereof. The agent could then determine the state with the maximal empowerment and then plan a sequence of actions to get to this state (Leu et al., 2013). This solution is, of course, infeasible in general.

Here, however, we will, in accordance with the latter principle, generally present the computed empowerment values as an empowerment map, because it gives an overview over what behavior would be preferred in either case. The map visualizes both the local gradients and the optima. The behaviors resulting from these maps would then be either an agent that acts in order to climb up the local gradient or to reach a global optimum.

3. EMPOWERMENT PERSPECTIVES

In this section, we outline how we can use the empowerment formalism to capture the essential aspects of the Three Laws. For this, we look at a system that contains both a human and a robotic agent. The Causal Bayesian Network from **Figure 2** shows the combined perception–action loop of both a human and a robot. There is a jointly accessible world R that both the human and robotic agent can perceive and act upon with their respective
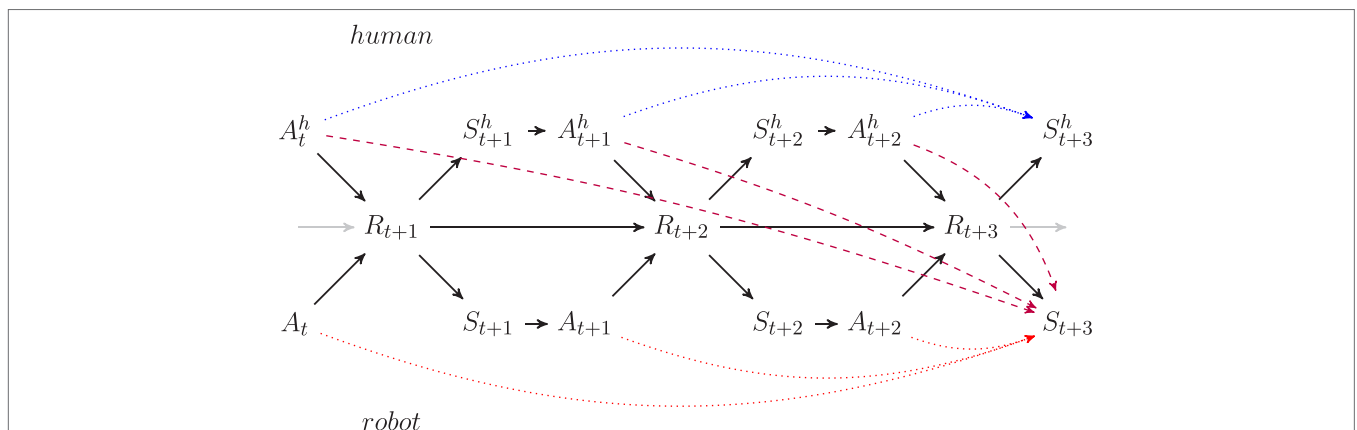

FIGURE 2 | The time-unrolled perception–action loop with two agents visualized as a Bayesian Network. S is the robot sensor, S^h the human sensor, A is the robot actuator, A^h is the human actuator, and R represents the rest of the system. The index t indicates the time at which the variable is considered. The arrows are causal connections of the Bayesian networks, the dotted and dashed line denote the three types of causal information flows relevant to the three types of empowerment discussed in the text. The red dotted arrows indicate the direction of the potential causal flow relevant for 3-step *robot empowerment*, the blue dotted arrows denotes *human empowerment*, and the dashed purple line indicates the human-to-robot *transfer empowerment*.

sensor and actuator variables. In such a system, it is possible to define different empowerment perspectives: here, we will look at *robot empowerment*, *human empowerment*, and *human-to-robot transfer empowerment*.

Robot empowerment is defined as the channel capacity with respect to the robot's actuators A and sensor S. This is the classical empowerment perspective. The robot choosing actions to maximize robot empowerment leads to behavior that preserves and enhances the robots ability to act and affect is surroundings. The propose, therefore, that robot empowerment can provide a heuristic for self-preservation and reliability.

Human Empowerment is similarly defined as the channel capacity with respect to the human's actuators A^h and sensors S^h. The difference here is that it is still the robot that chooses its own actions as to maximize the human's empowerment. This should create robot behavior aimed at preserving and enhancing the human's ability to act. This provides a heuristic that ultimately protects the human from environmental influences that would destroy the human or reduce its ability to act. It also creates behavior that aims to enhance the human's access and influence on the world and keeps the robot from hindering or harming the human directly. This heuristic somewhat corresponds to the First Law and provides a degree of safety.

Human-to-robot transfer empowerment is defined as the channel capacity from the human's actuators A^h to the robot's sensors S. This captures the potential causal flow from the human's actions to the world perceived by the robot. If the robot acts to maximize this value, it will maintain something we like to call operational proximity. It will keep the robot close to the human in a way where close does not necessarily mean physically close, but close so that the human can affect the robot. Furthermore, transfer empowerment can also be raised by the robot acting reliable, i.e., reacting in a predictable manner to the human's actions. This enhances the humans ability to act and would allow for the human to use certain actions to direct the robot. We propose, therefore, that this heuristic captures certain aspects of reliability and compliance, without directly reproducing the Second Law of robotics.

In combination these three heuristics should provide an operationalized motivation for robots to act in a way that reflects the sentiment behind the Three Laws of Robotics. The core of this idea was initially suggested by Salge et al. (2014b). Guckelsberger et al. (2016) have later used similar perspectives (they used transfer empowerment in the opposite direction) to generate companion behavior for a non-player character (NPC) in a dungeon crawler game.

3.1. Robot Empowerment

Robot empowerment is the potential causal information flow from the robot's actuators to the robot's sensors at a later point in time. In this perspective, the human is simply included in the external part of the perception–action loop of the robot. From the robot's perspective, all variables pertaining to the human are subsumed in R. The human's influence on the transition probabilities from A to S become relevant to it only as part of the robot's "Umwelt" and as such they are integrated into the robot's local forward model. Therefore, we expect robot empowerment behavior to be similar to what is observe in existing work on single agent

empowerment maximization (Klyubin et al., 2005, 2008; Salge et al., 2014c), where empowerment behavior aims to maintain the agent's freedom of operation and, indirectly, its survivability.

Typical empowerment-driven behavior can be seen, e.g., in control problems, where empowerment maximization balances a pendulum and a double pendulum, and also stabilizes a bicycle (Jung et al., 2011). We emphasize that in these examples no external reward needs to be specified, and empowerment derives directly from the intrinsic dynamics of the system. In other words, empowerment identifies the balancing positions as goals without an external specification of a goal, because these states offer the greatest degree of simultaneous control *and* predictability. Notably, empowerment has a tendency to drive the agent away from states where it would become inoperational, corresponding to a breakdown or "death" of an organism. Importantly, this does not require an external penalty for death, as breakdown or death are directly represented in the formalism via the vanishing of empowerment. Typically, a negative empowerment gradient can serve as an alert to the agent that it is in danger of moving toward destructive (or loss-of-control) states (Salge et al., 2014a). Guckelsberger and Salge (2016) summarize a lot of the self-preservation aspects of empowerment and argue that empowerment maximization can lead to autopoesis (Maturana and Varela, 1991).

More concretely relevant for the present argument is the work by Leu et al. (2013), pictured in **Figure 3**, where a physical robot uses the 2D-map of its environment to create an empowerment map to modulate its navigation. While the robot realized its primary objective to follow a human, it also tried to maintain its own empowerment, thereby avoiding to get stuck or to navigate into a tight passage which would reduce its ability to act. In the experiment, the reaction of the environment to the robot's actions is learnt using Gaussian process models. The human-following behavior itself is hard-coded and the human behavior itself is not covered by the model.

While the human-in-the-loop can be treated, from the robot's perspective, as just another part of the environment, we can modify our question and ask how the maximization of robot empowerment will affect the human or what kind of interactive behavior will result from this? Consider **Figure 4**, depicting a simplified model of the set-up in **Figure 3**. The figures show that, once we simulate the human behavior as part of the model, the robot's empowerment is drastically reduced around the human. First, in this particular scenario the robot's safety shutdown turns it off in proximity to the human, and its empowerment drops to zero. So the human can perform (move) actions that will obliterate the robot's empowerment. Other setups will have a different and possibly less drastic response, but in any case the presence of the human impacts the robot's empowerment. Notably, if the robot's resulting sensor state can be influenced by the human's actions, it is possible for the human to disturb the outcome of the robot's behavior. Depending on how well the robot can predict the human's action, the anticipated outcome, i.e., its sensor state at the time horizon, will be more or less noisy. Empowerment selectively avoids interactions with unpredictable agents (Salge et al., 2013) and generally noisy, unpredictable situations. It lowers the robot's control over its own environment and thereby its
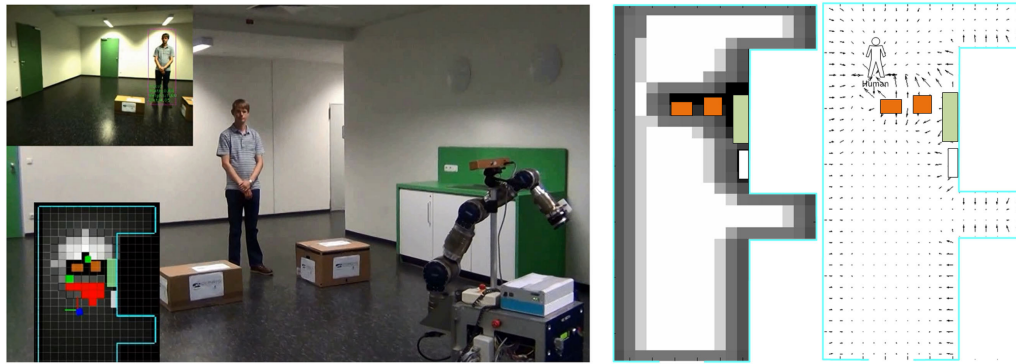
FIGURE 3 | Empowerment-driven robot follower as described in Leu et al. (2013). The robot tries to follow the human while trying to maintain its empowerment in the obstacled environment. By trying to maintain empowerment, the robot keeps away from obstacles and walls and, furthermore, avoids passing through the narrow passage separating it from the human because it would constrain its movement. The left subfigure shows the experimental setup, with a scaled empowerment map on the bottom left. The top left shows the robot's view, used for tracking the human. The right subfigure shows the room layout with empowerment map overlaid; right: gradient vector field of the scaled empowerment map, scaled by Euclidean distance to the human to induce following behavior.
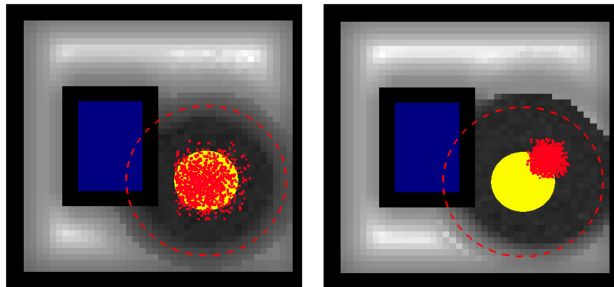

FIGURE 4 | Robot empowerment (in grayscale: dark—low, bright—high) dependent on the robot position. Obstacles in blue. Two different human (yellow circle) behavior models are considered. If the robot would be within the safety shutdown distance (red circle), it is not able to act. The red dots are the endpoints of the 2,000 random human action trajectories used for possible action predictions, the two graphs differ in the assumed distribution of human movement in the next three steps. If one compares the empowerment close to the safety shutdown distance, one can see that the assumed human behavior influences the estimated empowerment while all other aspects of the simulation remain unchanged.

sensor input; this loss of predictable control expresses itself in a loss of empowerment.

The effects of this can be seen in **Figure 4**, where the robot empowerment landscape is changed purely by a different human behavior model. On the right, we have a more accurate model of the predicted human behavior, which also assumes that the human moves toward the upper left. This moves the shutdown zone where the robot empowerment vanishes and at the same time also "sharpens" the contrast between the high- and low-empowered areas. The sharper contrast is a result of the reduced amount of noise being injected by the human into the robot's perception–action loop.

A reliably predictable human would also allow the robot to maintain a higher empowerment closer to the human if it would stay south-west of the human, basically opposite of the humans predicted movement direction. In general, obtaining a better human model can increase the robot's perceived model-based empowerment by reducing the human noise and by providing a better estimate which states in close proximity to the humans are less likely to become disempowering.

A related phenomenon was observed in a study by Guckelsberger et al. (2016), where three different empowerment perspectives were used to control a NPC companion in a NetHack-like dungeon crawler game. In one simulation, the player was able to shoot an arrow which would kill the NPC. This caused the empowerment-driven NPC agent to always avoid standing in the direction of the player. The NPC's world model assumed that all actions of the player were equally likely, and therefore, assumed that there is a chance that the player would kill the NPC. This led to a tendency to avoid the player. Guckelsberger et al. showed that this could be mitigated by a "trust" assumption, basically changing the player model as to assume the player to be benevolent to the NPC and would not perform an action that would lead to a loss of all NPC empowerment (i.e., kill the NPC). This made the NPC less inclined to flee from the player. Applied to a real robot, this trust assumption would make sure that the robot would focus on mitigating actual possible threats rather than having to hedge against malevolent or even just negligent human action ("friendly fire").

Better model acquisition is necessary for an agent to gage whether or not to avoid humans, depending on their cooperation, unpredictability, or antagonism. But empowerment maximization offers, by itself, not a direct incentive toward or away from human interaction. It is possible to imagine that the human could perform actions that would increase or preserve the robots or NPCs empowerment, such as feeding or repairing/healing it. This would then likely create a drive toward the human, if this is modeled in viable timescales. In absence of any specific possible benefits provided by the human, however, the agent driven by empowerment alone is not automatically drawn toward interacting with a human and is likely to drift away over time. This will be addressed in the other perspectives below.

In short, we propose to maximize robot empowerment to generate behavior by which the robot strives toward self-preservation.

Specifically, becoming inoperational corresponds to vanishing empowerment. In turn, having high empowerment means that the robot has a high influence on the world it can perceive, implying a high readiness to respond to a variety of challenges that might emerge. We propose this principle, namely, maximizing empowerment, as a generic measure as a plausible proxy for producing behavior in the spirit of the Third Law, as it will cause the robot to strive away from states where it expects to be destroyed or inoperational, and to strive toward states where it achieves the maximum potential effect. Robot empowerment maximization thus acts to some extent as a surrogate for a drive toward self-preservation.

3.2. Human Empowerment

We now turn to human empowerment. Human empowerment is defined in analogy to robot empowerment as the potential causal flow from the human's actuators to the human's sensors. The robot is now part of the external component of the perception–action loop of the human. Maximizing, or at least preserving human empowerment has similar effects to the previous case: keeping it at a high value implies maintaining the human's influence on the world and avoiding situations which would hinder or disable the human agent. A central difference to the previous case is that the human empowerment is made dependent on the *robot*; in other words, now the robot aims to maximize the empowerment of another agent rather than its own.

Figure 5 shows another continuous 2D scenario, this time involving a laser. The laser would block the human's movement, but is harmless to the robot and can be blocked by it. The grayscale coloring this time indicates the value of the human empowerment,
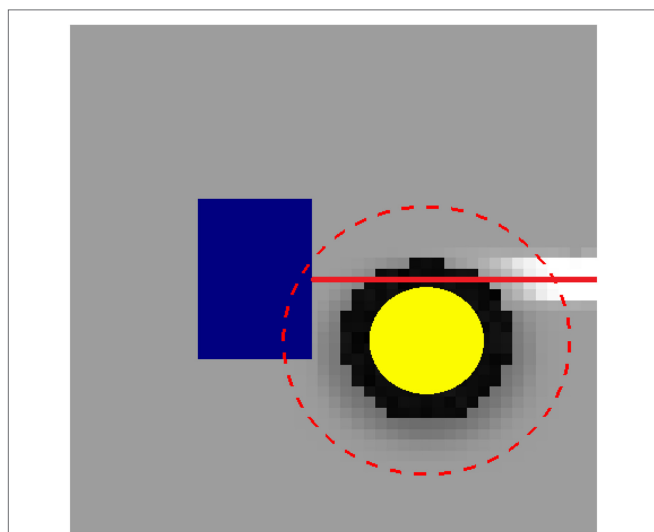

FIGURE 5 | Human empowerment (in grayscale: dark—low, bright—high) dependent on the robot's position. In this simulation, a laser (indicated by the red line) blocks the human's movement, but the laser can be occluded by the robot body. Thus, Human empowerment is the highest for robot positions toward the right wall where the robot blocks the laser and thereby allows the human a greater range of movement.

but still depending on the robot's position. Notably, the highest empowerment is achieved when the robot is in the white area. This is where the robot blocks the laser coming from the right side, permitting the human to move past the laser barrier. Driven by maximization of human empowerment, the robot would, therefore, prefer to move in front of the laser. Importantly, note that high human empowerment values for the robot in this area only become pronounced when the human is close to the laser barrier (as is the case in the figure, the human position denoted by the yellow circle). This means that the drive to block the laser only emerges at all when the human is actually in a position to pass it.

We can also see in **Figure 5** that positions in the area directly around the human produce low human empowerment. This is because the robot would partially block the humans movement if it were in these positions close to the human. Also note that the empowerment landscape for the robot being further away from the human is flat. Here, the robot is so far away that it does not interact in any way with the human's action–perception loop, and therefore, its exact position has no influence on the human's empowerment.

For another illustrative example of the effect of considering human empowerment-driven robots, consider the NPC (non-person character; autonomous, computer-controlled player in a video games) in the dungeon crawler game scenario from Guckelsberger and Salge (2016). Here, the NPC avoids standing directly next to the (human) player, as this would block the player's movement. If enemies are present that are predicted to shoot and kill the player, the NPC strives to kill these enemies to save the player. This can be combined with a maximization of the NPC's own empowerment. **Figure 6** shows a sequence of NPC actions that arise from combining the different heuristics. Here, the heuristics is a linear combination of the different empowerment types. As a result, the NPC will first save itself (as the human player has more health) and then it removes the two enemies. This complex behavior emerges from a greedy maximization of the linear combination of the different empowerment perspectives.

In this section, we considered human empowerment maximization (in the variant of being influenced by the artificial agent rather than the human). Using this variant as driver leads to a number of desirable behaviors. It prevents the robot from obstructing the human, for example, by getting too close or by interfering with the human's actions. Both would be noticeable in the human empowerment value, because they would either constrain accessibility to states around the human, or inject noise in the human's perception–action loop. In addition to that, the robot acts as to enhance or maintain the human's empowerment, through "proactive"-appearing activities, represented in above examples by removing a barrier from the environment or by neutralizing a threat that would destroy or maim the human or even just impede their freedom of movement. In this sense, human empowerment maximization can be plausibly interpreted as a driver for the agent toward protecting the human and supporting their agenda.

A number of caveats remain: to compute the human empowerment value, the robot not only needs a sufficient forward model but also needs to be able to identify the human agent in the environment, their possible actions, and how the human
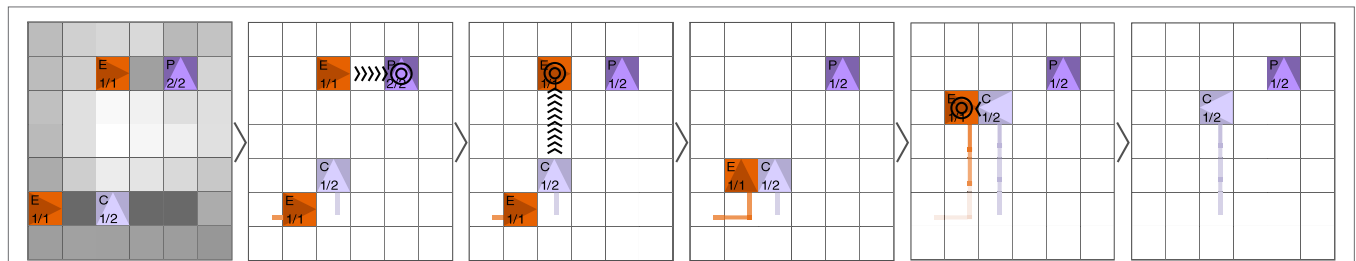
FIGURE 6 | Companion (C) and player (P), both purple, are threatened simultaneously by two enemies (E), both red. The images represent the successive moves: the companion escapes its own death, rescues the player, and finally defends itself. Left: combined heuristics for 2-step empowerment (see text), lighter colors indicating higher empowerment. Arrows indicate shooting. Figure taken from Guckelsberger and Salge (2016), reproduced with permission.

perceives the world via their sensors and what they are able to do with their actuators. This is not a trivial problem; however, it nevertheless has the advantage that it offers, in some ways, a "portable" and operational modeling route. While it depends on a sufficiently reliable algorithm for detecting humans and plausible, if strongly abstracted, models for human perception and actuation, once these are provided, the principle is applicable to a wide range of scenarios. The present proposal suggest possible routes toward an operational implementation of a "do not cause harm to a human" and a "do not permit harm to be caused to a human" principle, provided one can endow the artificial agent with a—what could loosely be termed—"proto-empathetic" perspective of the human's situation.

Another critical limitation to the applicability of the formalism is the time horizon, which is the central free parameter in the empowerment computation. While a robot driven by human empowerment maximization might stop a bullet, or a fall into a pit, it would need to extend its time horizon massively to account for things that would be undesirable for the human in the short-term, but are advantageous in the long run. To illustrate, consider the analogy from human lawmaking, where freedom to act on the short scale is curtailed in an effort to limit long-term damage (e.g., in environmental policies). The principle as discussed in this section is, therefore, best suited for interactions that have to avoid obstruction or interference by a robot in the short term and with immediate consequences. That being said, nothing in the formalism prevents one—in principle—to be able to account for long-term effects. To do so in practice will require extending the methods to deal with longer time-scales and levels of hierarchies.

3.3. Transfer Empowerment

As the third and last variant, we consider transfer empowerment. Transfer empowerment is defined as the potential causal information flow from the actions of *one* agent to the sensors of *another*. One of the motivations for its development was to counter the lack of the two previous perspectives to provide an incentive that keeps the robot/companion from drifting away from the human. The aim was to add an incentive for the robot to remain at the human's service. This is achieved by requiring that the human's actions can influence the sensory input of the robot.

Figure 7 shows the human-to-robot (HtR) transfer empowerment, i.e., we consider the human movement as the empowerment-inducing action set and the empowerment-relevant sensor
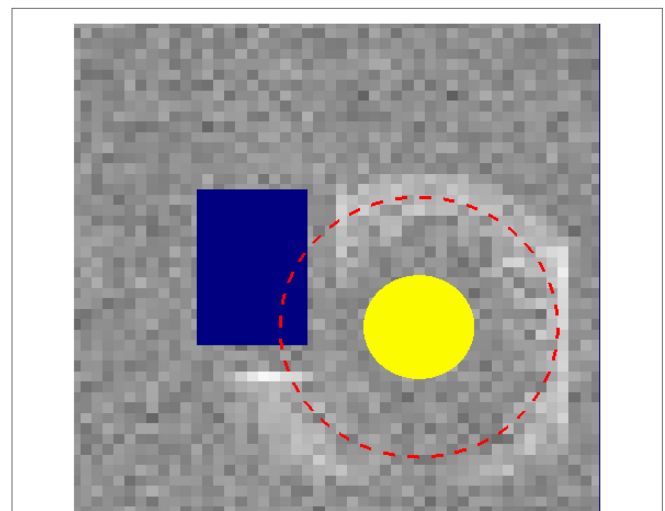

FIGURE 7 | A visualization of the human-to-robot (HtR) transfer empowerment dependent on robot position (dark—low, light—high). This simulation shows a slightly elevated human-to-robot transfer empowerment around the human agent (yellow) at the shutdown distance (red circle). This is because the human can move toward or away from the robot, thereby having the potential to stop the robot. This creates a potential causal flow from the human's actions to the robot's sensors, which in this case measure the robot position. Here, the physical proximity of the human allows it to directly influence the robot's state.

input is given by the robot positions. Here, we employ the safety shutdown mechanism at the fixed distance as a simple illustrative proxy for other, potentially more complex influences that the human may have on the robot. In our case, the shutdown mechanism allows the human to selectively disable the robot, by moving toward or away from it. This creates a direct causal influence from the human to the robot. Consequently, this generates a ring of higher HtR transfer empowerment around the human. In general, any form of influence of the human on the perception of the robot will produce a modulation of the HtR transfer empowerment landscape. By maximizing this value, the robot would try to remain in a domain where the human can affect it.

This effect can alternatively be obtained by the analogous robot-to-human transfer empowerment; this was demonstrated in Guckelsberger and Salge (2016), see **Figure 8**. Here, the companion tries to maximize the causal influence its actions have on
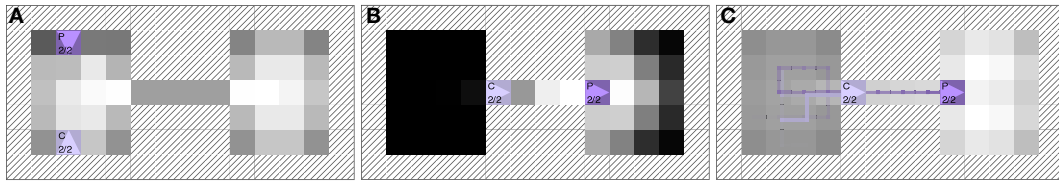
FIGURE 8 | Two room scenario from Guckelsberger and Salge (2016), reproduced with permission. **(A)** Companion empowerment, n = 2. **(B)** Companion-player transfer empowerment, n = 2. **(C)** Coupled empowerment with movement trace, n = 2. In the last subfigure **(C)**, the companion agent C is driven by a combination of its own empowerment and transfer empowerment from companion to player (values shown in grayscale, bright—high, dark—low). If only companion empowerment is driving the agent, as in subfigure **(A)**, then the companion does not enter the corridor, as its narrowness will lower the agent's own empowerment. With the addition of transfer empowerment, however, the companion begins to maintain operational proximity, and thus follows the player also through the narrow corridor. This can be seen in the movement trace of both player and agent.
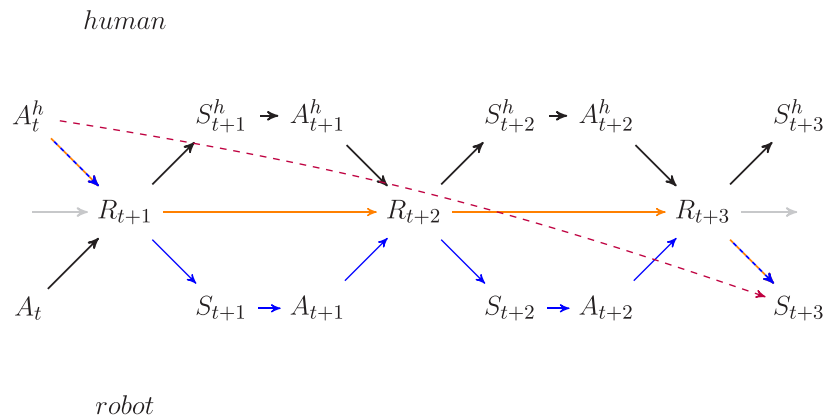

FIGURE 9 | The time-unrolled perception–action loop of two agents, colored to visualize the different pathways potential causal flow can be realized. The red dashed line indicate the causal flow from the human actuator A_t^h at time t to the robot's sensor variable S_{t+3} at time $t + 3$, which contributes to human-to-robot transfer empowerment. This potential causal flow can be realized by direct human influence on the environment R (one exemplary path shown in blue). Alternatively, the human can influence the environment R, and the robot can then perceive this change, react to it by choosing an appropriate action A and thereby influence its perceived environment itself. One exemplary path is shown in orange. The dashed arrows are those used by both pathways.

the human player's input state, which also causes it to remain close to the player. This particular variant of empowerment helped to overcome the persistent problem that the companion would not follow the player through narrow corridors, as they were strongly constraining its empowerment.

So, in regard to creating player-following behavior, both directions of transfer empowerment seem to be suitable. But, looking at the causal Bayesian network representation in **Figure 9**, we can see that there are in principle two different ways how the human's action can affect the robot's sensors (and vice versa). One way is for the human's actions to directly change the world R in a way that will be perceived by the robot. The other way is for the human to also affect the world R, but then for the robot to detect this change via its own sensors, and react based on this input, changing parts of R itself. In the second case, the information flows through the internal part of the perception–action loop of the robot—in the first case it does not. An example of the second case can be seen in **Figure 10**, where the robot moves in the same direction as the human, if there is a direct line of sight between the two agents. This results in a high transfer empowerment in those areas where the human can be seen by the robot. Obtaining this

high transfer empowerment requires the robot to react reliably to certain human actions.

The distinction between the two pathways for transfer empowerment, directly through the environment and through the internal part of the other agent's perception–action loop, also provides us with reasons to prefer human-to-robot transfer empowerment over transfer empowerment in the other direction. In human-to-robot transfer empowerment, the internal pathway is through the agent; so the robot can consider adjusting its behavior, i.e., the way it responds with actions to sensor inputs, in order to increase the transfer empowerment. In the robot-to-human transfer empowerment, the internal pathway is through the human, which the robot cannot optimize and the human should ideally not be burdened with optimizing. So, if one seeks to elicit a reliable reaction of the robot to the human's action, then human-to-robot transfer empowerment should be the quantity to optimize.

Another difference between the two directions of transfer empowerment becomes evident when we compare **Figures 4** and **7**. Both show the same simulation, but depict HtR- and robot empowerment, respectively. The area with higher HtR transfer
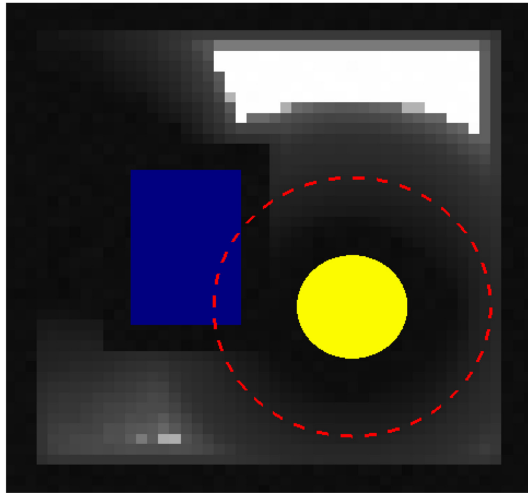
FIGURE 10 | A visualization of the human-to-robot (HtR) transfer empowerment dependent on robot position (dark—low, light—high). In this scenario, the robot will mirror the human's movement if it has a direct line of sight. This creates a high amount of potential causal flow through the robot where the latter sees the human. It results in comparatively high transfer empowerment in those areas where the robot has both a direct line of sight to the human (yellow) and is not shut down by close proximity to the human. The highest transfer empowerment is attained at a distance, but in an area where the robot can see and react to the human; with this, the robot provides the human with operational proximity, i.e., the ability to influence the robot's resulting state.

empowerment is exactly the same area where the robot empowerment around the human begins to drop. This is because while the human here gains control over the robot's position, the latter loses this very control. Controllability of a specific shared variable in the environment is a limited resource, and if one agent has full control of it, the other agent consequently has none. This is another reason why the use of transfer empowerment is usually preferable in the human-to-robot direction, namely, as to not provide an incentive for the robot to take control away from the human.

The different scenarios we looked at also illustrate the idea of "operational proximity" that transfer empowerment captures. The influence of one agent on another does not necessarily depend on physical proximity, but rather on both agents' embodiment, here in the form of their actions and sensor perceptions. While in one scenario the human could stop the robot by physical proximity, in the other they could direct the robot along their line of sight. In the dungeon example, the companion needs proximity to directly affect the player by blocking or shooting it, but one could also instead imagine a situation where the NPC would push a button far away or block a laser somewhere else to affect the environment of the player. Maximizing transfer empowerment tries to attain this operational rather than physical proximity. In turn, operational proximity acts as a necessary precondition for any interaction and coordination between the agents. To interact, one agent has to be able to perceive the changes of the world induced by the other agent. Vanishing transfer empowerment would mean that not even this basic level of interaction is possible.

Furthermore, HtR transfer empowerment maximization also creates an incentive to reliably react to the human actions. In detail, this means increasing the transfer empowerment further by allowing for some potential causal flow through the *internal* part of the robot's perception–action loop. Note that for the empowerment calculation, it does not matter how precisely robot actions are matched to human actions; all that counts is that by consistently responding in the same way the robot effectively extends the human's empowerment in the world. The robot reacting to the human expands the influence of the latter on the world, because the human's actions are amplified by the robot's actions. An additional effect is that, if a robot reliably reacts to the human's actions, the human can learn this relationship and use its own actions as proto-gestures to control the robot. On the one hand, this is still far removed from giving explicit verbal orders to the robot, as described in the Second Law. On the other hand, such a reliable reaction of the robot to human actions would permit humans to "learn" a command language consisting of certain behaviors and gestures that would then cause the robot to respond in a desired way.

Summarizing the section, while the maximization of transfer empowerment does not precisely capture the Second Law, it creates operational proximity between the human and the robot, and thereby the basis for further interaction; together with the enhancement of human empowerment, it sets the foundation for the human to have the maximum amount of options available. Furthermore, if the robot behavior itself is included in the computation of transfer empowerment to be optimized, then this would provide an additional route to amplify the human's actions in the world, namely by virtue of manipulating the robot via actions and proto-gestures which make use of an implicitly learnt understanding of the internal control of the robot.

4. DISCUSSION

The core aim of this article was to suggest three empowerment perspectives and to propose that these allow—in principle—for a formalization and operationalization of ideas roughly corresponding to the Three Laws of Robotics (not in order): the self-preservation of a robot, the protection of the robot's human partner, and the robot supporting/expanding the human's operational capabilities. Empowerment endows a state space cum transition dynamic with a generic pseudo-utility function that serves as a rich preference landscape without requiring an explicit, externally defined reward structure; on the other hand, where desired, it can be combined with explicit task-dependent rewards. Empowerment can be used as both a generic and intrinsic value function. It serves not only as a warning indicator that one is approaching the boundaries of the viability domain (i.e., being close to areas of imminent breakdown/destruction) but also imbues the interior of the viability domain with additional preference structure in advance of any task-specific utility. This, together with the properties outlined in Section 2.1 about intrinsic motivations, suggests empowerment as a useful driving principle for an embodied robot.

For the practical application of the presented formalism to real and more complex robot–human interaction scenarios, still

a number of issues, such as computability and model acquisition need to be addressed. The following discussion outlines suggestions on how some of these challenges can be overcome and the problems still to be solved to deploy the presented heuristics on real-world scenarios.

4.1. Computability

Extending the idea presented here from simple abstract models into the domain of practical robotics immediately raises the questions of computability. In the classical empowerment formalism, computation time scales dramatically with an increase in sensor and actuator states and with an extension of the temporal horizon. In discrete domains, previous work has demonstrated a number of ways as to how to speed up the computation: one being the *impoverished empowerment* approach by Anthony et al. (2011). Here, one only considers the most "meaningful" action sequences to generate the empowerment, pruning away the others, and then building up longer action sequences by extending only those meaningful action sequences.

A simple alternative option is to just sample a subset of all action sequences and compute empowerment based on this sample to get a heuristic estimate for the actual value (Salge et al., 2014a). This approach only works effectively in a system with discrete states and deterministic dynamics that does not spread out too much.

Earlier, we mentioned a fast approximation method for continuous variables (Salge et al., 2012). This assumes that the local transition can be sufficiently well approximated by a Gaussian channel. Another recent and fast approximation for Gaussian channels uses variance propagation methods, implemented using neural networks (Karl et al., 2015). Another promising approach utilizing neural networks is presented by Mohamed and Jimenez Rezende (2015) and provides not only considerable speed-ups but also has the advantage of working without an explicit forward model.

With the establishment of empowerment as a viable and useful intrinsic motivation driver for artificial agents in a battery of proof-of-principle scenarios over the last years, it has become evident that it is well warranted to invest effort into improving and speeding up empowerment computation for realistically sized scenarios. The previous list of methods shows that a promising range of approaches to speed up empowerment computation already exists and that such approximations may well be viable. We have, therefore, grounds to believe that future work will find even better ways to scale empowerment computation up, thus rendering it more suitable to deployment on practically relevant robotic systems.

4.2. Model Acquisition

Traditional methods for empowerment calculation crucially require the agent to have an interventional forward model (Salge et al., 2012). So far, we have sidestepped this question of model acquisition, mostly because the acquisition of the local interventional model to compute empowerment can be treated as a separate problem which can be solved in different ways by existing methods (Dearden and Demiris, 2005; Nguyen-Tuong and Peters, 2011).

First, one would learn the local causal dynamics of agent–world interaction; from this one can then compute empowerment which, as a second step, provides an intrinsic reward or pseudo-utility function which is associated with the different world states distinguishable to the agent. As example, previous work in the continuous domain demonstrated how Gaussian Process learners (Rasmussen and Williams, 2006) can learn the local dynamics of a system on the fly while empowerment based on these incrementally improved, experience-based models can then be used to control inverted pendulum models (Jung et al., 2011) or even a physical robot (Leu et al., 2013).

Whether the forward model is prespecified or learnt during the run, empowerment will generally drive the agent toward states with more options. However, if trained during the run of the agent, the model will in general also include uncertainty on the outcome of actions. Such uncertainty "devalues" any options available in this state and will lead to a reduction of empowerment. This reduction is irrespective of whether the uncertainty is due to "objective" noise in the environment and unpredictability (e.g., due to another agent) or due to internal model errors stemming from insufficient training. From the point of view of empowerment both effects are equivalent.

The first class of uncertainty (environmental uncertainty) will tend to drive the agent away from noisy or unpredictable areas to comparatively more predictable ones if the available options are otherwise equivalent, or reduce the value of richer option sets when they can be only unpredictably invoked. The second class of uncertainty (model uncertainty) will—in the initial phase of the training—cause empowerment to devalue states where the model cannot resolve the available options. This has consequences for a purely empowerment-driven exploration of rich, but non-obvious interaction patterns; a prominent candidate for such a scenario would be learning the behavior of other agents, as long as they are comparatively reliable.

The intertwining of learning and empowerment-driven behavior can thus be expected to produce a number of meta-effects on top of the already discussed dynamics. This could range from exhibiting a very specific type of exploratory behavior; moreover, such agents might initially be averse to encountering complex novel dynamics and other agents. On the other hand, by modulating the learning process and experience of an agent as well as its sensory resolution or "scope of attention" depending on the situation, one could guide the agent toward developing the desired sensitivities.

Such a process would, in a way, be reminiscent of the socialization of animals and humans to ensure that they develop an appropriate sense for the social dynamics of the world they live in. This, together with the earlier discussion in this paper, invites the hypothesis that, to be confident of the safety of an autonomous robot the following is essential: not only does this machine need to be "other-aware," but if that "other-awareness" is to be learnt while enjoying to a large extent the level of autonomy that an intrinsic motivation model provides, it will be essential for the machine to undergo a suitably organized socialization process.

As a technical note, we remark that the type of model required for empowerment computation only needs to relate the actions

and current sensor states to the expected subsequent sensor states and does not require the complete world mechanics. In fact, empowerment can be based on the general, but purely intrinsic *Predictive State Representations* (PSR) formalism (Singh et al., 2004; Anthony et al., 2013). This makes it suitable for application to recent robotic approaches (Maye and Engel, 2013) interested in the idea of sensorimotor contingencies (O'Regan and Noë, 2001); this work considers the understanding of the world to be built up by immediate interaction with it and learning which actions change and which do not change the agent's perception. This offers a route to deploy empowerment already on very simple robots with the aim of gradually building an "understanding" of the world from the bottom up.

4.3. Partial Sensor Empowerment

One central property of empowerment is that the formalism remains practically unchanged for different incarnations of robots or agents and has very few parameters; the time horizon being the only one for discrete empowerment. However, there is one, less obvious, "parameter" which has only been briefly discussed in the previous literature (Salge et al., 2014c) and only hinted at in the last section: the question of sensor and actuator selection in the model.

By considering only certain sensor variables, one can reduce the state space and speed up computation immensely. However, one also influences the outcome of the computation by basically assigning which distinctions between states of the world should be considered relevant. An agent with positional sensors will only care about mobility, while an agent with visual sensors might also care about being in a state with different reachable views. In the simplest models, we often just assume that the agent perceives the whole world. In biological examples, we can lean on the idea of evolutionary adaptation (i.e., Jeffery (2005)), arguing that sensors that would register states irrelevant to the agent would disappear, leaving only relevant sensors. Meanwhile, on a robot, we usually have a generous selection of sensors, and we might not consider all of them to be relevant. However, for parsimonious design, or in imitation of a hypothesized principle of parsimony in biology (Laughlin et al., 1998), it might be opportune to limit oneself to selecting essentially those sensors associated with capabilities worth preserving. Similarly, when considering the human empowerment, modeling the appropriate human sensors will make a big difference to which operational capacities of the human, and which forms of influence on the world, will be protected by the robot.

This becomes even more of an issue when considering transfer empowerment which is basically an example of using partial sensor selection to focus on relevant properties. If both the human and the robot could fully sense the environment, then the human empowerment and the human-to-robot transfer empowerment would be identical, as both the human and the robot sensors would capture exactly the same information. In the continuous 2D example presented in this paper, we considered the human's sensors to only capture the human's position, and the robot sensors to only capture the robot's position, so their perceptions would be distinct. In the NPC AI example by Guckelsberger and Salge (2016), the player and companion have sensors that

perceive the world around it within a certain distance, and always relative to their own position. That is, they detect the content of the field directly to the east of the player, rather than specify the field of interest by a coordinate value such as 3.1.

If we extended the human's sensors to the extent that the human could sense at least everything that the robot can sense, then the sensors relevant for transfer empowerment would be a subset of human empowerment. We could then just compute human empowerment and capture both potential causal flows at the same time. But by splitting the sensor variables, we basically compute partial sensor empowerment, once for the sensors pertaining to the human state and once with a selection of sensors pertaining to the robot state. In a real world scenario the embodied perspective of both the human and the robot usually lead to different perceptions of the world. Both have limited sensors and as a result there is a natural distinction between their respective sensor inputs. When using simulated environment, on the other hand, it is often easy to give all agents access to the whole world state. In this case it becomes necessary to limit the sensors of the different agents to introduce this split, before one is able to differentiate between the two different heuristics. This separation of human and human-to-robot empowerment allows for the prioritization of human empowerment over transfer empowerment. In general, we would expect one bit of human empowerment to be more valuable than one bit of human-to-robot empowerment and, therefore, aim to retain this distinction.

4.4. Combination of the Heuristics

Whenever different variants of some evaluation function exist which cover different aspects of a phenomenon and these aspects are combined, one is left with the question how to weight them against each other. In the present case, this would mean balancing the three types of empowerment. The analogy with the Three Laws might suggest a clear hierarchy, where one would first maximize human empowerment and then only consider the other heuristics. However, given that one can always expect some minimal non-trivial gradient to exist in the first, such a lexicographic ordering would basically lead to only maximizing human empowerment above all else, completely overriding the other measures.

On the other hand, going back to **Figure 6**, we saw that the companion, when faced with a threat to itself *and* the player, choose to first save itself from death, while permitting minor damage to the player, and only then proceeded to remove the threat to the player. While this clearly violates the strict hierarchy of the Three Laws, such a course of action might well be in the rational interest of the player, since it might be worthwhile to trade the minor loss of a life point for still having a companion (all this, of course, presumes that the agent's forward model is correct that the enemy shooting the human once will not seriously damage the latter; but such dilemmata are also present in mission-critical human decision-making under uncertainty).

This conceptual tension is also present in the original Three Laws. Consider a gedankenexperiment where a robot is faced with two options: (A) inflicting minor harm, such as a scratch, on a single human or (B) by avoiding that, permitting the destruction

of all (perfectly peaceful) robots on earth. In a strict interpretation, the Laws would dictate to chose option (B), but we might be inclined to consider that there must be some amount of harm so negligible that (A) would seem the better option than cause a regression to a robotic "Stone Age."

But how could we capture this insight with our three previously developed heuristics? We can, of course, consider a straightforward weighted sum of the three heuristics, defining some trade-off between the three values in the usual manner. But this approach inevitably raises the question whether there would be some distinct non-arbitrary trade-off.

The previous analogy is instructive. The problem is that option (B), the destruction of all robots, would create a lot of more significant problems further down the line than the single scratch of option A. It would result in the loss of all robots able to carry out the human commands, and there would be fewer robots to protect and save humans in the future (they would have to be rebuilt, absorbing significant productive work, before—if at all—reaching original levels).

Both of these problems are reflected in human empowerment on longer timescales. In fact, we would suggest that all three heuristics, and actually also the original Three Laws, reflect the idea that one core reason why humans build and program robots is actually to increase their very own empowerment, their very own options for the future. We already argued that transfer empowerment, the second heuristic, extends human empowerment further into the world, because the robot amplifies the human's actions and their impact on the world. Similarly, robots that preserve themselves, as by following the third heuristic, make sure that they preserve or extend the human's empowerment further.

The first heuristic is already directly about human empowerment maximization itself. So, in essence, the two other heuristics, robot empowerment and transfer empowerment, can be seen as a form of meta-heuristics for ultimate human empowerment maximization. We conjecture that both the behaviors of the second and the third heuristic might emerge once one maximizes the human empowerment with a sufficiently long temporal horizon. For example, the robot could realize, with a good enough model of the future, that it needs to keep itself functional in order to prevent harm to the human in the future. So, basically, we hypothesize that the Second and Third Law might manifest themselves as a short term proxy for a suitable longer term optimization of human empowerment. If so, it may be that this would help define a natural trade-off: in our example, the robot might calculate that, by preventing destruction of all robots on earth, at cost of inflicting a small scratch on a human, it would prevent many more and worse injuries of humans in the future by the thus rescued robots.

4.5. Multi-Agent Empowerment

One final remark: in this paper we have not considered true multi-agent empowerment. The reason for this is subtle: empowerment so far is usually computed as an open-loop channel capacity. The future (potential) action sequences considered for empowerment are basically executed without reacting to the changing sensor states inside the time horizon. In other words, empowerment is computed as the channel capacity between fixed-length "open-loop" action sequences and the future sensor observation. In choosing these action sequences, intermediate sensor observations *are not* taken into account. Thus the agent does not react to particular developments in the environment while probing the potential future actions.

This makes it impossible to formally account for instantaneously reacting to another agent's actions during the computation of the potential futures, since this model selects actions only at the beginning and then only evaluates how the will affect the world at the end of the action run. However, we showed earlier that transfer empowerment can be massively enhanced by reacting to the human's actions. This indicates strongly that it would be important to model empowerment with "reactive" action sequences, i.e., empowerment where action sequences are expressed in closed-loop form and which instantaneously react to other agents (or even changes in the environment) while still *inside* the time horizon of the probed futures.

These various aspects of the implementation of empowerment indicate a number of strategies to render it a useful tool to operationalize the Three Laws in a transparent way. Furthermore, they also may offer a pathway demonstrating how also other classes of intrinsic motivation measures might be adapted to achieve the fusion of the desirable autonomy of robots with the requirements of the Three Laws.

AUTHOR CONTRIBUTIONS

CS and DP developed the original idea. CS wrote and ran the simulations and drafted the paper. DP advised during the development of the simulations and co-wrote the paper.

FUNDING

REFERENCES

Anderson, S. L. (2008). Asimov's "three laws of robotics" and machine metaethics. *Ai Soc.* 22, 477–493. doi:10.1007/s00146-007-0094-5

Anthony, T., Polani, D., and Nehaniv, C. (2013). General self-motivation and strategy identification: case studies based on Sokoban and Pac-Man. *IEEE Trans. Comput. Intell. AI Games* PP, 1. doi:10.1109/TCIAIG.2013.2295372

Anthony, T., Polani, D., and Nehaniv, C. L. (2011). "Impoverished empowerment: 'meaningful' action sequence generation through bandwidth limitation," in *Advances in Artificial Life. Darwin Meets von Neumann – 10th European Conference, ECAL 2009, Budapest, Hungary, September 13–16, 2009, Revised Selected Papers, Part II, Volume 5778 of Lecture Notes in Computer Science*, eds G. Kampis, I. Karsai, and E. Szathmáry (Budapest, Hungary: Springer), 294–301.

Arimoto, S. (1972). An algorithm for computing the capacity of arbitrary discrete memoryless channels. *IEEE Trans. Info. Theory* 18, 14–20. doi:10.1109/TIT.1972.1054753

Asimov, I. (1942). Runaround. *Astound. Sci. Fiction* 29, 94–103.

Asimov, I. (1981). The three laws. *Compute* 18, 18.

Ay, N., Bertschinger, N., Der, R., Güttler, F., and Olbrich, E. (2008). Predictive information and explorative behavior of autonomous robots. *Eur. Phys. J. B Condens. Matter Complex Syst.* 63, 329–339. doi:10.1140/epjb/e2008-00175-0

Ay, N., and Polani, D. (2008). Information flows in causal networks. *Adv. Complex Syst.* 11, 17–41. doi:10.1142/S0219525908001465

Blahut, R. (1972). Computation of channel capacity and rate-distortion functions. *IEEE Trans. Info. Theory* 18, 460–473. doi:10.1109/TIT.1972.1054855

Boden, M., Bryson, J., Caldwell, D., Dautenhahn, K., Edwards, L., Kember, S., et al. (2011). "Principles of robotics," in *The United Kingdoms Engineering and Physical Sciences Research Council (EPSRC)* (Web Publication). Available at: https://www.epsrc.ac.uk/research/ourportfolio/themes/engineering/activities/principlesofrobotics/

Brooks, R. (1986). A robust layered control system for a mobile robot. *IEEE J. Robot. Auto.* 2, 14–23. doi:10.1109/JRA.1986.1087032

Coradeschi, S., Loutfi, A., and Wrede, B. (2013). A short review of symbol grounding in robotic and intelligent systems. *Künstliche Intell.* 27, 129–136. doi:10.1007/s13218-013-0247-2

Dearden, A., and Demiris, Y. (2005). "Learning forward models for robots," in *Proceedings of the 19th International Joint Conference on Artificial Intelligence* (Edinburgh: Morgan Kaufmann Publishers Inc.), 1440–1445.

Dennett, D. C. (1984). "Cognitive wheels: the frame problem of AI," in *Minds, Machines and Evolution*, ed. C. Hookway (Cambridge, MA: Cambridge University Press), 129–150.

Der, R., Steinmetz, U., and Pasemann, F. (1999). *Homeokinesis: A New Principle to Back up Evolution with Learning*. Leipzig: Max-Planck-Inst. für Mathematik in den Naturwiss.

Guckelsberger, C., and Salge, C. (2016). "Does empowerment maximisation allow for enactive artificial agents?" in *Proceedings of the 15th International Conference on the Synthesis and Simulation of Living Systems (ALIFE'16)*, Cancun.

Guckelsberger, C., Salge, C., and Colton, S. (2016). "Intrinsically motivated general companion npcs via coupled empowerment maximisation," in *IEEE Conference on Computational Intelligence and Games (CIG)* (Santorini: IEEE).

Hasslacher, B., and Tilden, M. W. (1995). Living machines. *Rob. Auton. Syst.* 15, 143–169. doi:10.1016/0921-8890(95)00019-C

Jeffery, W. R. (2005). Adaptive evolution of eye degeneration in the Mexican blind cavefish. *J. Hered.* 96, 185–196. doi:10.1093/jhered/esi028

Jung, T., Polani, D., and Stone, P. (2011). Empowerment for continuous agent environment systems. *Adapt. Behav.* 19, 16. doi:10.1177/1059712310392389

Karl, M., Bayer, J., and van der Smagt, P. (2015). Efficient empowerment. arXiv preprint arXiv:1509.08455. Available at: https://arxiv.org/abs/1509.08455

Klyubin, A., Polani, D., and Nehaniv, C. (2005). "Empowerment: a universal agent-centric measure of control," in *The 2005 IEEE Congress on Evolutionary Computation*, Vol. 1 (Edinburgh: IEEE), 128–135.

Klyubin, A., Polani, D., and Nehaniv, C. (2008). Keep your options open: an information-based driving principle for sensorimotor systems. *PLoS ONE* 3:e4018. doi:10.1371/journal.pone.0004018

Laughlin, S. B., de Ruyter van Steveninck, R. R., and Anderson, J. C. (1998). The metabolic cost of neural information. *Nat. Neurosci.* 1, 36–41. doi:10.1038/236

Leu, A., Ristic-Durrant, D., Slavnic, S., Glackin, C., Salge, C., Polani, D., et al. (2013). "Corbys cognitive control architecture for robotic follower," in *IEEE/SICE International Symposium on System Integration (SII)* (Taipei: IEEE), 394–399.

Maturana, H. R., and Varela, F. J. (1991). *Autopoiesis and Cognition: The Realization of the Living*, Vol. 42. Springer.

Maye, A., and Engel, A. K. (2013). Extending sensorimotor contingency theory: prediction, planning, and action generation. *Adapt. Behav.* 21, 423–436. doi:10.1177/1059712313497975

McCauley, L. (2007). Ai Armageddon and the three laws of robotics. *Ethics Inf. Technol.* 9, 153–164. doi:10.1007/s10676-007-9138-2

Mohamed, S., and Jimenez Rezende, D. (2015). "Variational information maximisation for intrinsically motivated reinforcement learning," in *Advances in Neural Information Processing Systems 28*, eds C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, R. Garnett, and R. Garnett (Montreal: Curran & Associates Inc.), 2116–2124.

Murphy, R. R., and Woods, D. D. (2009). Beyond Asimov: the three laws of responsible robotics. *IEEE Intell. Syst* 24, 14–20. doi:10.1109/MIS.2009.69

Nguyen-Tuong, D., and Peters, J. (2011). Model learning for robot control: a survey. *Cogn. Process.* 12, 319–340. doi:10.1007/s10339-011-0404-1

Oesterreich, R. (1979). *Entwicklung eines Konzepts der objectiven Kontrolle und Kontrollkompetenz. Ein handlungstheoretischer Ansatz*. Ph.D. thesis, Technische Universität, Berlin.

O'Regan, J. K., and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.* 24, 939–973. doi:10.1017/S0140525X01000115

Oudeyer, P.-Y., and Kaplan, F. (2007). What is intrinsic motivation? A typology of computational approaches. *Front. Neurorobot.* 1:6. doi:10.3389/neuro.12.006.2007

Oudeyer, P.-Y., Kaplan, F., and Hafner, V. (2007). Intrinsic motivation systems for autonomous mental development. *IEEE Trans. Evol. Comput.* 11, 265–286. doi:10.1109/TEVC.2006.890271

Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. Cambridge, UK: Cambridge University Press.

Pfeifer, R., and Bongard, J. (2006). *How the Body Shapes the Way We Think: A New View of Intelligence*. Cambridge, MA: MIT Press.

Rasmussen, C., and Williams, C. (2006). *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press, 1.

Ryan, R. M., and Deci, E. L. (2000). Intrinsic and extrinsic motivations: classic definitions and new directions. *Contemp. Educ. Psychol.* 25, 54–67. doi:10.1006/ceps.1999.1020

Salge, C., Glackin, C., and Polani, D. (2012). Approximation of empowerment in the continuous domain. *Adv. Complex Syst.* 16, 1250079. doi:10.1142/S0219525912500798

Salge, C., Glackin, C., and Polani, D. (2013). "Empowerment and state-dependent noise-an intrinsic motivation for avoiding unpredictable agents," in *Advances in Artificial Life, ECAL*, Vol. 12, 118–125. Available at: https://mitpress.mit.edu/sites/default/files/titles/content/ecal13/ch018.html

Salge, C., Glackin, C., and Polani, D. (2014a). Changing the environment based on empowerment as intrinsic motivation. *Entropy* 16, 2789. doi:10.3390/e16052789

Salge, C., Glackin, C., and Polani, D. (2014b). "Empowerment: a route towards the three laws of robotics," in *Seventh International Workshop on Guided Self-Organization*. Freiburg. Available at: http://ml.informatik.uni-freiburg.de/events/gso14/index

Salge, C., Glackin, C., and Polani, D. (2014c). "Empowerment – an introduction," in *Guided Self-Organization: Inception* (Springer). Available at: https://link.springer.com/chapter/10.1007/978-3-642-53734-9_4?no-access=true

Schmidhuber, J. (1991). "Curious model-building control systems," in *IEEE International Joint Conference on Neural Networks* (Singapore: IEEE), 1458–1463.

Seligman, M. E. (1975). *Helplessness: On Depression, Development, and Death*. New York, NY: WH Freeman/Times Books/Henry Holt & Co.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 623–656. doi:10.1002/j.1538-7305.1948.tb00917.x

Singh, S., James, M. R., and Rudary, M. R. (2004). "Predictive state representations: a new theory for modeling dynamical systems," in *Proceedings of the Twentieth Conference on Uncertainty in Artificial Intelligence (UAI)*, Banff, 512–519.

Steels, L. (2004). "The autotelic principle," in *Embodied Artificial Intelligence: International Seminar, Dagstuhl Castle, Germany, July 7–11, 2003. Revised Papers*, eds F. Iida, R. Pfeifer, L. Steels, and Y. Kuniyoshi (Berlin, Heidelberg: Springer), 231–242.

Sutton, R. S., and Barto, A. G. (1998). *Reinforcement Learning*. Cambridge, MA: MIT Press.

Trendafilov, D., and Murray-Smith, R. (2013). "Information-theoretic characterization of uncertainty in manual control," in *2013 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Manchester, 4913–4918.

von Foerster, H. (2003). *Disorder/Order: Discovery or Invention? Understanding Understanding Book Subtitle Essays on Cybernetics and Cognition*. New York: Springer, 273–282.

von Uexküll, J. (1909). *Umwelt und Innenwelt der Tiere*. Berlin: Springer.

Wissner-Gross, A., and Freer, C. (2013). Causal entropic forces. *Phys. Rev. Lett.* 110, 168702. doi:10.1103/PhysRevLett.110.168702

Ziemke, T., and Sharkey, N. E. (2001). A stroll through the worlds of robots and animals: applying Jakob von Uexküll's theory of meaning to adaptive robots and artificial life. *Semiotica* 134, 701–746. doi:10.1515/semi.2001.050