Disentangling Reafferent Effects by Doing Nothing

Benedict Wilkins, Kostas Stathis

Department of Computer Science, Royal Holloway University of London
benrjw@gmail.com, kostas.stathis@rhul.ac.uk

Abstract

An agent's ability to distinguish between sensory effects that are self-caused, and those that are not, is instrumental in the achievement of its goals. This ability is thought to be central to a variety of functions in biological organisms, from perceptual stabilisation and accurate motor control, to higher level cognitive functions such as planning, mirroring and the sense of agency. Although many of these functions are well studied in AI, this important distinction is rarely made explicit and the focus tends to be on the associational relationship between action and sensory effect or success. Toward the development of more general agents, we develop a framework that enables agents to disentangle self-caused and externally caused sensory effects. Informed by relevant models and experiments in robotics, and in the biological and cognitive sciences, we demonstrate the general applicability of this framework through an extensive experimental evaluation over three different environments.

Introduction

An agent's ability to distinguish between sensory effects that are self-caused, and those that are not, is instrumental in the achievement of its goals. A seminal work (von Holst 1954) on this distinction in biological agents coined the terms *reafference* and *exafference* to mean: the parts of an observation that are caused by the agent's own action, and the parts of an observation that are caused by external conditions or events respectively. The subtlety of the distinction is highlighted by Von Holst:

> If I shake the branch of a tree, various receptors of my skin and joints produce a reafference, but if I place my hand on a branch shaken by the wind, the stimuli of the same receptors produce an exafference.

The distinction has played a central role in developing theories to explain a broad range of physiological phenomena (von Holst 1954; Blakemore, Wolpert, and Frith 2000; Wolpert and Flanagan 2001; Medendorp 2011; Fukutomi and Carlson 2020).

In some of the earliest experiments performed by Mittelstaedt and Von Holst (von Holst 1954) it was demonstrated that reafference plays an important role in the modulation of sensory signals. If a fly is placed inside a vertically striped rotating cylinder, in an attempt to stabilise itself the fly will rotate its body to compensate for the perceived motion. This optomotor-reflex, is not observed when the fly moves of its own accord in a stationary cylinder, even though the sensory consequences are essentially equivalent. If the head of the fly is rotated 180 degrees, effectively switching the position of its eyes, and the cylinder is again rotated, as one might expect the fly will rotate itself in the opposite direction. However when again moving of its own accord it will overcompensate and begin to spin. The inversion of the modulating reafferent signal leads the optomotor-reflex to correct in the wrong direction. The mechanism that underpins these observations was made more precise in subsequent work (Miall and Wolpert 1996) in which a comparative theory of reafference was introduced, see Fig. 1. The efference copy (or corollary discharge (Sperry 1950)), a copy of an internal outward motor signal or *action*, along with a forward model (Kawato 1999; Wolpert, Ghahramani, and Jordan 1995; Wolpert and Flanagan 2001) estimates the reafferent sensory consequences. The estimate is compared with subsequent sensory data, and any error is attributed to exafference. For the fly, the exafferent passing of stripes across its optical field is a signal to correct its motion if knocked off course or blown by the wind.

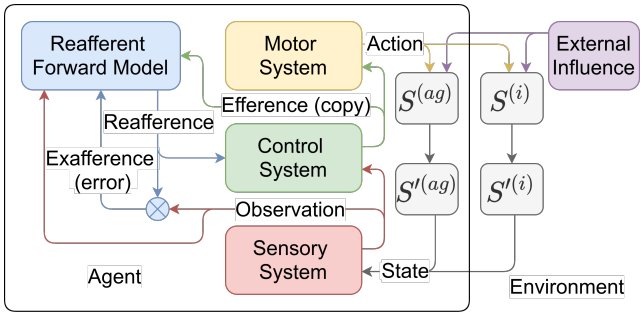

Figure 1: Comparative view of reafference. The forward model estimates reafferent effects from the efference copy and the agent's observation. Reafference is compared with subsequent observations to determine exafference.

Although there is substantial experimental evidence for the theory (Straka, Simmers, and Chagnaud 2018), it does

not inherently explain how the forward model comes to produce good estimates of the reafferent signal. A biological agent experiences changes in their motor system over their lifetime, for example, through growth, disease or misfortune in the case of the fly. Experiments with humans suggests that the forward model is to some extent adaptive (Wilke, Synofzik, and Lindner 2013), and that reafferent estimates are learned through experience. If reafference is to be learned through experience, an important question is raised: if an agent only ever experiences a mixture of reafference and exafference as the total afferent signal, then how does the forward model come to model only the reafferent signals? In other words, how is an agent able to draw the distinction between sensory effects that are self-caused and those that are externally-caused.

The aim of this work is to investigate this fundamental question. Specifically, we are interested in mechanisms that can disentangle reafference from exafference in ways that preserve causal relations between action and sensory effects. After all, it is these causal relations that provide an agent with the information that is most widely applicable to the task at hand. As one might expect, reafference, although not referred to as such, turns out to be central to many important problems in AI, ranging from credit assignment and planning to localisation and stabilisation in robotics.

Drawing on the long history of reafference in AI, and inspired by experiments with biological agents, our aim is to develop a formalisation of reafference that is representative of a general causal perspective. We believe such a formalisation to be a step in the right direction for many of the important open problems in AI and that it could have wider conceptual implications in related fields.

Our core contributions lie in our formalisation of reafference as a causal estimand and an algorithm that allows an agent to disentangle reafference and exafference from experience. The key ingredient in our approach, both practically and theoretically, is the act of *doing nothing*. Rather than attributing errors to external influence, external influences are modelled explicitly and any error is considered simply as model error which should be minimised. Crucially, the act of doing nothing allows an agent to draw the desired distinction by reasoning counterfactually about the effects of its actions. In an attempt to ground our work, which has both conceptual and practical components, we are informed by relevant concepts and experiments performed in the biological and cognitive sciences, using them as a basis for our own experiments with artificial agents.

Background

An agent is situated in an environment where it can take action. The agent's actions have some effect on the state of the environment, and therefore on what the agent observes. These effects may be initially unknown to the agent as they are in our setting, or may be known (to the extent that they can be inferred) in advance. The environment in which the agent is situated may evolve without the agent taking action, for example through the action of another agent, or by some other process. Where self-caused effects are known in advance, the problem of disentanglement is deferred to the source of the reafferent model, which for artificial agents is the developer.

In early work on reasoning about action in AI, the causal relationships between actions and their effects are assumed known and are represented as a program expressed in a logical form. This program acts as the agent's reafferent forward model with which it can infer self-caused effects. In early work in situation calculus, for example (McCarthy 1963) and its derivatives, the environment evolves with statements of the form $do(A, S)$ where A is an action (e.g. $move(x, y)$) and S is a situation or state of the world. The agent can query the logic program with an action to obtain the logical consequences. The agent also might generate a plan by instead querying with a goal state, where the plan is generated by reasoning about the consequences of action. Using a language like PDDL for example, a similar approach can be taken for planning in robotics. The agent is provided with an *a priori* forward model that is derived from the structure and properties of its body. This model is a kind of causal model and is based on our knowledge of physics; we have again taken time to disentangle the relevant causal relationships ourselves.

Although planning in these settings is still a challenging problem and useful in many applications, we cannot always rely on knowing, or being able to specify the causal relations that are required for building a reafferent model for the agent. It is not always clear how each aspect of the agent's observation is related to its action; vision is a particularly difficult example. This has led to the development of methods that instead try to learn or discover the relevant relationships automatically.

As discussed previously, this is what biological agents do in one form or another through their experience, or perhaps through their evolution in the case of a fly. One view of how biological agents might draw the distinction is that the predictability of a sensory signal determines whether it is reafferent or exafferent, with the reafferent signal being the more predictable. It is *easy* to estimate the reafferent signal produced by shaking a tree branch, but more difficult to predict the observed exafference that is due to the wind. The implementation of this view is typically quite crude, and rather ends up as a model of the association between action and sensory effect. One possible implementation is to remove external influence altogether, making exafference *unpredictable* only to the extent that a model has not previously seen these influences and is therefore bad at predicting their effects.

In (Schroder-Schetelig, Manoonpong, and Worgotter 2010), a bipedal robot learns to walk. The forward model is trained when the robot is situated on a flat surface, the robot is later tested on sloped surfaces. The robot is successful in stabilising itself in the new sloped environment using its forward model and exafferent error signal. This approach has again deferred the problem of disentanglement, leaving it up to us to determine a suitable training environment. Although the agent is now free to learn the effects of its actions, it is generally difficult to create an environment that is free of external influence. In this instance, the learned forward model suffers from bias. It has not disentangled the effect

on its observation that is due to gravity. Clearly the gravitational effect is not due to the agent's action and should be considered exafferent. If we wanted to send this robot to Mars it would fail to stabilise since the forward model is working with the measure of Earth's gravity. One might argue that gravity need not be modelled as exafference if we are not on an interplanetary mission since it is constant on Earth. Nevertheless, there is a conceptual issue to address and that is that the gravitational effect is not *caused* by the agent's action, and that there are other similar variables of interest that may be difficult to control for. The assumption that underpins this experiment can be found in a number of other works (Bechtle, Schillaci, and Hafner 2016; Schillaci et al. 2016). The essential issue with the approach is that the choice of environment determines what effects are considered reafferent.

Going further still, in order to maximise return and therefore solve the task it has been given, a reinforcement learning (RL) agent must model the long-term effects of its action and typically does so via its value function. It is not clear to what extent model-free RL agents learn reafference. They are able to exploit and maximise return, but likely work with an associative model of an action's effect on observation (or return) rather than a causal one. What is clear is that RL agents are attentive to only those aspects relevant for maximising return (Lapuschkin et al. 2019). A similar phenomenon is seen in the other learning paradigms, most clearly in supervised learning (Geirhos et al. 2020). This selective associational modelling of reafferent signals by RL agents leads to less robust policies, worse generalisation performance, and exacerbates problems with learning long term dependencies between action and return. If for example, early in training an agent finds that particular aspects of (or effects on) its observation lead to reward in the short term, it may neglect to model those aspects that turn out to be relevant for obtaining more reward long term.

In addressing some of these open problems, a number of works have tried to provide agents with a means to better learn the effects of their actions, often by introducing notions that are related to causality, such as counterfactuals (Buesing et al. 2018; Mesnard et al. 2021) or *imagination* (Schrittwieser et al. 2020). They have also been a key motivation for works most closely related to ours (Bellemare, Veness, and Bowling 2012; Corcoll and Vicente 2020).

Decisions, Actions & Interventions

To act is to bring about change in an environment. One popular formalisation of action comes from decision theory, where an agent's decision is represented as a variable A whose outcome is an action. The effects of the action are determined by the relationship between A and the variables that represent the state of the environment. A decision is made by an agent given its observations and beliefs about the world, which are themselves variables, in pursuit of a goal.

In Pearl's conception of causal inference (Judea Pearl 2000), actions are instead represented as interventions, that is, changes to the underlying relationships between state variables, and not as outcomes of decision variables. In their simplest form, they fix the value of a variable, for example $do(X = x)$, where X is some state variable and x is a value.

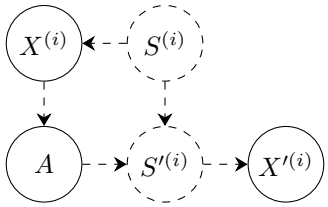

Figure 2: Dashed nodes indicate unobserved variables. Dashed edges show the direction of causation as in a standard causal graph. They additionally indicate that some edges may be missing between variables in the corresponding collections. For example, a particular variable that is present in the agent's observation $X^{(i)}$ may not have any baring on its decision. Similarly, the action the agent decides upon may not affect every part of the state $S'^{(i)}$.

To determine causal relations in practice we find ourselves intervening on variables that look very similar to what might be called decision variables in decision theory. To be concrete, consider the following prevalent introductory example: a new treatment T is to be tested for its effectiveness in combating a particular disease. The effect of T on patient health Y is to be estimated to determine whether the drug is suitable for wider use. Although T is typically referred to as the treatment variable, it in fact represents a medical practitioner's decision to give ($T = 1$), or not give treatment ($T = 0$). The intervention is therefore a modification of the decision making mechanism; $do(T = 1)$ will fix the value of the decision to be give treatment, regardless of patient heath, age etc. Of course, causal inference is more general than this, one can intervene on any observed state variable, not just those that look like decision variables.

For the purposes of our work, an action is the outcome of a decision variable. An intervention is a modification of the mechanism that determines the outcome of the decision variable, $do(A = a)$ sets the agent's decision to the action a without regard for the agent's observations or beliefs. This setup allows us to formalise reafference as *the changes in an agent's observation that are due to changes in the decision variable A.*

Formalising Reafference

Formally, we consider the following stochastic control process. The environment transitions according the distribution $\mathcal{T}(S_{t+1} = s_{t+1}|S_t = s_t, A_t = a_t)$. $\mathcal{T}$ gives the probability of an environment transitioning to a particular state s_{t+1} given the current state s_t and the agent's action a_t. An agent observes x_t, which is derived from the state according to the distribution $\mathcal{O}(X_t = x_t|S_t = s_t)$. Actions are selected by the agent according to a policy $\pi(A_t = a_t|X_t = x_t)$ or by intervention $do(A_t = a_t)$. The time index t is useful for determining the direction of causal relations, however going forward we drop it as we are generally interested in the effect of an action on a particular observation x. We treat each triple (X_t, A_t, X_{t+1}) in isolation, with X' denoting the

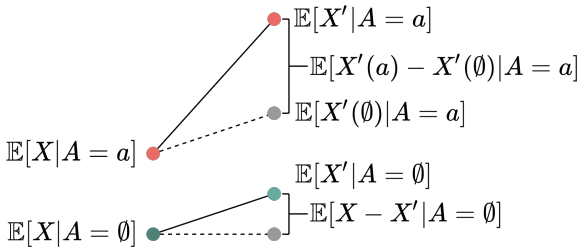

Figure 3: Difference in differences to get the average reafferent effect. The individual reafferent effect can be found similarly by conditioning on the current observation x, with realisations of X' being the possible observations after x.

possible observations after a specific observation x. The degenerate random variable X is used to denote the current observation. The potential outcome of the agent's next observation taking action a is denoted as $X'(a)$. The causal graph that we work with is presented in Fig. 2.

Doing Nothing

In the treatment problem presented previously, T is a binary variable whose outcome is determined by the practitioners decision, give treatment, do not, or more generally, act, do not. Not acting, or *doing nothing* turns out to be crucial in determining reafference. This action, which we denote $\emptyset$ and refer to as the *null*-action, sets a baseline that allows the agent to reason *If I do nothing, there are only exafferent effects.* For many environments the choice of $\emptyset$ is quite natural, for example, as the action with least (ideally zero) expenditure of energy. Or, in videos games, where the action is commonly referred to as *noop* (no operation) and represents an absence of input from the player. The decision variable A may be continuous or discrete, but requires such a null-action to be explicitly chosen.

Reafferent Effects

The Average Causal Effect (ACE), sometimes referred to as the Average Treatment Effect (ATE), is a common causal estimand of interest in many settings. The average reafferent effect (ARE) turns out to be well captured by this estimand and corresponds to the ACE of action on observation.

$$\delta(a) = \mathbb{E}[\bar{X}'(a)] - \mathbb{E}[\bar{X}'(\emptyset)]$$

The ARE is taken over all observations, i.e. over all time steps, as indicated by the $\bar{\square}$, see Fig. 3. As an illustrative example, consider the following Structural Causal Model (SCM):

$$\begin{aligned} A &:= Bernoulli(p_a) \\ Z &:= Bernoulli(p_z) \qquad & Z' &:= Bernoulli(p_z) \\ Y &:= \mathcal{N}(0,1) & Y' &:= (A * Z) + Y \end{aligned}$$

It depicts an agent taking an action A to move into an empty space $Z = 1$, the space may already be occupied $Z = 0$. The agent observes $X = (Y, Z)$ where Y is the agent's current position. $Z = 1$ is a precondition for the successful execution of the action A, otherwise Y does not change (i.e. $Y' = Y$). By inspection it can be seen that $\delta(1) = (p_z, 0)$. To get the *individual* reafferent effect one can condition on a specific observation:

$$\delta(a|X = x) = \mathbb{E}[X'(a)|X = x] - \mathbb{E}[X'(\emptyset)|X = x]$$

In the example above, we need only to condition on Z as the reafferent effect is invariant to Y. Again by inspection, $\delta(1|Z = 0) = (0, 0)$ and $\delta(1|Z = 1) = (1, 0)$, corresponding to the unsuccessful/successful act of moving into the filled/empty space respectively. Here we are computing the reafferent effect for each observation, that is, for every observed value of Z and Y. This is possible because we have access to the SCM, if the quantity is to be estimated additional assumptions are required, specifically, that the estimator is representative of the SCM.

Assumptions

We make a number of assumptions that are standard in the causal inference literature: faithfulness, consistency, positivity, unconfoundedness, parallel trends and time independence. The following assumptions warrant some clarification:

Parallel trends The exafferent effect is the same regardless of the action taken for a particular observation (see Fig. 3).

Time independence The reafferent effect does not change with time. This assumption is broken if, for example, there is an unobserved state variable that interacts with time and with the action to produce its effect, such as age in biological agents.

Unconfoundedness There are no unknown confounders in the formalisation presented, the environment state contains all variables that might influence the next state (other than the agent's action). Although not all variables are observed by the agent, only those that the agent does observe have immediate causal relationships to its action, in other words, an agent makes a decision based only on what it observes. As such, conditioning on the observation X will eliminate any backdoor paths.

Uncounfoundedness can be broken in an alternative formulation in which actions are also dependant on unobserved state variables. Although it seems counter intuitive that actions could be decided based on something that is not observed, it is helpful to remember that, at least for biological agents, taking action is not an instantaneous process. The line between agent and environment is not as clear as it is for artificial agents. A may be influenced by the state of the agent's body, not all of which forms part of X. If the value of A is *measured* after this happens then S can be said to directly influence A. Another source of confounding might come from an agent's beliefs, since these are derived from previous observations, unless they are conditioned upon, backdoor paths may be present. To avoid these confounding issues, we assume agents take action based only on the current observation, and that actions are not influenced by S. These assumptions are reflected in the causal graph presented in Fig. 2 but may not be realistic for more complex (biological) settings.

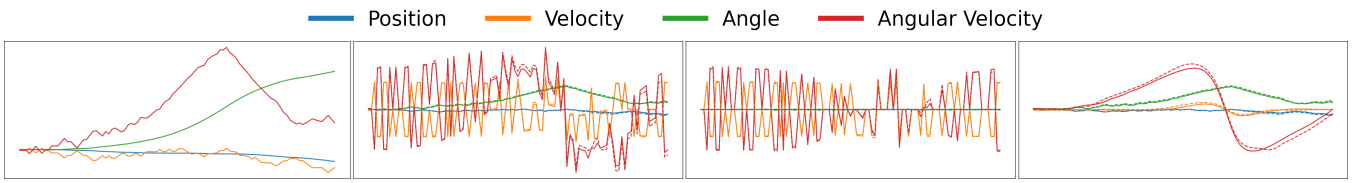

Figure 4: Disentangling reafference and exafference in Cartpole. Observations and (estimated) effects are shown over time, with the agent taking an action at each step. Graphs show (a) observation, (b) total effects, (c) reafferent effects, and (d) exafferent effects. Estimated effects are shown as dotted lines. An action is an instantaneous force applied to the cart which has an instantaneous effect on the carts (angular) velocity. Actions do not have an instantaneous effect on the carts position or angle.

Algorithm 1: Estimating Effects via SGD

Objective function $\mathcal{L}$ (MSE); Model f_θ; Policy π;
Environments $env^{(i)}$;
Initial observations $x^{(i)} \sim env^{(i)}$;
while *stopping criteria not met* **do**
  Sample actions $a^{(i)} \sim \pi(x^{(i)})$;
  Take actions $x'^{(i)} \sim env^{(i)}(a^{(i)})$;
  Ground truth (total) effect $\delta = \mathbf{x}' - \mathbf{x}$;
  Estimate exafference $\hat{\delta}_\emptyset = f_\theta(\mathbf{x}, \emptyset)$;
  Estimate reafference $\hat{\delta}_a = f_\theta(\mathbf{x}, \mathbf{a}) - \hat{\delta}_\emptyset$;
  Gradient $\nabla_\theta \mathcal{L}(\hat{\delta}_a + \hat{\delta}_\emptyset, \delta; \theta)$;
  Apply Gradient Update $\theta \leftarrow \theta - \eta \nabla_\theta$;
  $x \leftarrow x'$;
end

Estimating Reafference

Using the formalism presented in the previous sections, we develop an algorithm that trains a forward model to disentangle and estimate reafferent and exafferent effects. The algorithm is presented in Alg. 1.

An agent should learn the reafferent forward model as it gathers experience, this avoids issues with storing large amounts of observations for later use. A common trick that was devised in the deep reinforcement learning literature is to sample observations from multiple environments (Mnih et al. 2016), this ensures that the elements of the mini-batch are closer to i.i.d., a replay buffer (Lin 1992; Schaul et al. 2016) might also be used to break temporal correlation.

Assuming the use of a neural network f_θ with an MSE objective, the network will learn both the reafferent and exafferent effects. Each mini-batch is a collection of experience triples $(\mathbf{x}, \mathbf{a}, \mathbf{x}')$ each is a vector of observations/actions collected from the environments. The algorithm is estimating the *individual* reafferent effect for each observation/action. $f_\theta(x, \emptyset)$ is estimating the effect $\delta(\emptyset | X = x)$. $f_\theta(x, a)$ is estimating the total effect $\mathbb{E}[X' | A = a, X = x] - \mathbb{E}[X | A = a, X = x]$. The reafferent estimand $\delta(a | X = x)$ which is estimated as $f_\theta(x, a) - f_\theta(x, \emptyset)$ is expanded below.

$$\big[\mathbb{E}[X' | A = a, X = x] - \mathbb{E}[X | A = a, X = x]\big] - \big[\mathbb{E}[X'(\emptyset) | A = a, X = x] - \mathbb{E}[X(\emptyset) | A = a, X = x]\big]$$

It is assumed that the observed and counterfactual expectations of the current observation are equal, which here is trivially true since X is observed. The counterfactual quantity is estimated by the model by extrapolating from observed instances of doing nothing. The reafferent effect is 0 when $a = \emptyset$ since the same model is used to compute both observed and counterfactual quantities. The estimated total effects for both a and $\emptyset$ are compared with the ground truth to obtain the loss. To estimate the ARE the model can be provided with only the action as input. This may introduce confounding since X is no longer conditioned upon. To remedy, the agent can perform a randomised trial, removing the dependence on X and any potential backdoor paths by using a suitably random policy. The ARE can also be estimated by averaging over all individual effects, however the estimate may be biased by the policy.

Experimental Evaluation

Alg. 1 is applied to three different environments. To demonstrate the effectiveness and general applicability of our approach, we perform a number of experiments. Each showcases, or has a parallel with, an important concept or experiment performed in related fields. They are described in the relevant section below. Training and model details for each experiment, as well as additional experiments are presented in the supplementary material. All code and data is publicly available[1].

(i) Cartpole The Cartpole environment is a simple physical system with a cart that moves along a horizontal axis and a pole that should be balanced on top. This version of the environment has three actions $[-\beta, 0, \beta]$, each applies a horizontal force of magnitude β to the cart, with 0 applying no force (null-action). The agent observes the cart's position and velocity, and the pole's angle and angular velocity. Exafferent effects are due to gravitational acceleration, or velocities produced by previous actions. Reafferent effects are changes in the next observation that are due to the force applied to the cart by the agent.

Experiment (i.1) demonstrates that by doing nothing the agent can properly recover both reafferent and exafferent effects - including the constant gravitational effect. This is in contrast to the bipedal robot example (Schroder-Schetelig,

[1] https://github.com/BenedictWilkins/disentangling-reafference

Figure 5: Disentangling in the Atari Freeway environment (left) and Artificial Ape environment (right). Images show (a) current observation x, (b) ground truth total effect $x' - x$, (c) predicted reafferent effect, and (d) predicted exafferent effect. Effects are scaled $[-1, 1] \rightarrow [0, 1]$. See text for discussion.

Figure 6: Disentangling in Artificial Ape with congruent reafference and exafference. (a) observation (b) ground truth total effect (c) estimated total effect (d) estimated reafferent effect (e) estimated exafferent effect. In this example the agent and platform have rotated in opposite directions, leading to a cancellation in the total effect.

Manoonpong, and Worgotter 2010) in which the gravitational effect was absorbed by the forward model as exafference. Results for (i.1) are shown in Fig. 4. Actions have an instantaneous effect only on the velocities and not on position or angle, this is an environment implementation detail that is reflected in the result.

(ii) Atari Freeway In this environment the agent (a chicken) is attempting to cross a road while cars drive past. The agent, can move forward, backward, or stay where it is (null-action). The agent is presented with visual observations (images) and should disentangle the reafference, movement of its body, from exafference, cars driving past. The aim of this experiment is to show that our approach can separate the *body* of an agent from the rest of the environment. Results are presented in Fig. 5 (left).

Here body is meant loosely as the pixel representation of the agent where actions have local effects. In this instance, the agent does not explicitly recognise its body as an independent entity as we have given it no capacity to do so. It is however able to create a clean separation between reafference and exafference, where reafference happens to correspond to effects local to its body. The ability to distinguish body from environment is an important first step in developing embodied agents, in this instance for localisation, and for solving problems such as mirroring. A deeper exploration of reafference in these contexts is needed, but is currently beyond the scope of this work.

(iii) Artificial Ape The neural mechanisms that underpin reafference, and in particular the comparative theory have been investigated in numerous works. In one study, an ape was placed on a rotating platform, restrained but with some freedom to move its head. The authors studied vestibular neuron signals and found that there was indeed a distinction made between passive and active movement of the head (Cullen 2004; Roy and Cullen 2004).

In the spirit of these works, the Artificial Ape environment places an agent into a 3D scene with a collection of moving cubes (Wilkins and Stathis 2022). The agent can rotate its view left and right or maintain its current view (null-action). The movement of the cubes is independent of the agent's action and so is exafferent. The agent stands on a platform that is rotated randomly, rotating the agent's perspective with it. The goal here is two fold: (iii.1) to distinguish exafference on the visual system due to the movement of the cubes from the reafferent perspective shifts, and (iii.2) to distinguish the congruent exafferent perspective shifts from the reafferent perspective shifts. Results for (iii.1) are presented in Fig. 5 (right). The agent's action is to rotate its view by a small angle, leading to the highlighted vertical edges seen in the predicted reafferent effect. The chequered cubes in the scene are moving up/down relative to the view leading to the highlighted horizontal edges seen in the predicted exafferent effect. In this experiment the platform has been removed as it does not play a role.

Results for (iii.2) are presented in Fig. 6. To aid interpretation by reducing aleatoric uncertainty, the colour of the platform is an indicator of the future direction of motion of the platform. An analogue in the experiment with apes might be the whirr of the platform motor, or feeling of acceleration in the rest of the body. This introduces some incongruency to the effects, however the vast majority of variables are congruent. A more detailed analysis with truly congruent effects is presented in the supplementary material. The result shows a cancellation effect in the signals that is similar to what is observed in the original experiment when the apes head and platform move in opposite directions. The result suggests that our approach is viable as even in the case of congruence the agent is able to properly estimate the effects.

Discussion

Average Reafferent Effects (ARE) Our focus has been on estimating the *individual* reafferent effects, that is, the reafferent effect for a particular observation for some action. The primary reason for this is that reafference tends not to be same for all observations, this renders the ARE essentially useless. Consider for example, averaging over the visual effects in Freeway. If the reafferent effect is similar over all observations, or over subsets of observations, then the ARE

might be more useful to the agent. If in Artificial Ape instead of a visual observation the agent observed angular velocity then estimating the ARE might be appropriate.

Multi-step Reafference Reafference as defined thus far considers only single-step effects, that is, the effect of the agent's action on the next observation. In practice, effects may be extended in time, be delayed, or we may want to model the effect of an action over multiple time steps. Each case is essentially asking *what effect does action a_t have on observation x_{t+n}?*. To estimate multi-step effects, the model should compute counterfactual estimates for all intermediate actions during training. Then at test time take $n-1$ consecutive null-actions for comparison. An exploration of this is left as future work.

Gaps in Biological Reafference In some of the earliest work on reafference, Helmholtz noted that if one presses gently on their eye the world appears to move, however remains stationary when the eye is moved by the extraocular muscles (von Helmholtz and Southall 1924). This suggests that there is some reafferent modulation of the signal in the latter case that keeps the world stationary whenever our eyes saccade. It might also suggest a gap in biological reafference, since in the first instance the eye movement is also self-caused, just by a different motor mechanism and yet it is treated as exafferent. Our approach to modelling reafference has no such gaps. This might be an indicator that the mechanism behind biological reafference differs in some important way, or just that evolution has found shortcuts in cases where modelling such effects is really not necessary.

Higher Cognitive Functions Although the current motivation for learning reafference is for use in problems such as stabilisation, it has otherwise appeared in investigations of higher-level cognitive functions. For example in mirroring (Blakemore and Frith 2005; Rajmohan and Mohandas 2007) (theory of mind), the sense of agency (Haggard 2017) and the early development of *self* (Lewis 2012; Jékely, Godfrey-Smith, and Keijzer 2021), although there is some debate surrounding the extent of its role (Zaadnoordijk, Besold, and Hunnius 2019). While many of these functions are still out of reach in AI, it is clear that reafference plays an important role. Our hope is that by developing biologically inspired agents, such as those that can distinguish self-caused from externally-caused sensory effects, we gain some understanding of these problems. To use mirroring as an example, knowing the effects of one's actions seems crucial in being able to imitate another. Further, for one agent to recognise another as an agent with its own beliefs and intentions, it seems similarly crucial (Gallese and Goldman 1998).

Related Work

Contingency Awareness Contingency awareness (Watson 1966), a close conceptual relative of reafference is investigated in (Bellemare, Veness, and Bowling 2012). The term *contingent regions* was coined to mean the region of an observation that is affected by an agent's action. From a causal perspective, (Corcoll and Vicente 2020) defines a measure similar to that used to determine contingent regions. These measures are similar to our work in that action effects are compared counterfactually to determine a causal relation. However, they do not determine the causal *extent* of the relation. It is noted in (Corcoll and Vicente 2020) that a special *do-nothing* would not work well for estimating effects, arguing that doing nothing still has an effect on the observation (or at least the return). We believe this to be a conceptual oversight. While it is true that there is an effect on the observation, this effect should be ascribed to environmental influence. If the null-action is taken there is by our definition no edge from A to X' in the causal graph. This baseline is what allows us to determine the extent of the causal relation between the other actions and the agent's observation.

Associational Approaches The following approaches take advantage of strong regularisation or implicit model biases to perform some kind of disentanglement. The general idea is to condition on actions and inspect the internals of a model, or use salience maps, to determine the controllable regions of the observation. (Choi et al. 2019) takes advantage of spatial attention mechanisms and trains an inverse-dynamics model, (Yang et al. 2019) uses an action-conditioned beta-VAE, similarly (Zhong, Schwing, and Peng 2020) uses an action-information bottle neck with strong regularisation and (Oh et al. 2015) learns action-conditioned dynamics models. These works differ from ours in that they learn associational relations between action and observation. Additionally, measures of the relation are strongly subject to hyper-parameter choices and are biased by exafferent effects.

Controllable Factors of Variation One line of work (Thomas et al. 2018; Bengio et al. 2017; Sawada 2018) defines and makes use of *selectivity* as a measure of what they call *independent controllable factors of variation*. These factors correspond to aspects of the environment that are *controllable* independently of other aspects, for example, the chicken in the Freeway environment. There are parallels with our work in that the changes in these factors would for the most part be represented as reafferent effects. However, rather than effects, they are more abstract latent representations of what can be independently *done* in an environment. In Cartpole for example, the cart and pole as a whole would be modelled as a factor, and the chicken in Freeway as another. The factors in Artificial Ape are less clear.

Conclusions & Future Work

In this work we have investigated reafference in the context of artificial agents and AI. We formalised reafference as a causal estimand that can be estimated by a counterfactual comparison of doing nothing and doing something. This gives an agent the means to distinguish between self-caused and externally-caused sensory effects. By drawing links between reafference in AI and related fields, through our approach we have taken steps towards developing more general biologically inspired agents. Future work includes exploring some of the higher-level cognitive functions that are related to reafference. For example, by integrating our approach into a planning system, reinforcement learning agent, or a mirroring mechanism.

Acknowledgements

We would like to extend our appreciation to François Torregrossa and Chris Watkins for their valuable feedback. This work was funded by the Techne AHRC Doctoral Training Partnership.

References

Bechtle, S.; Schillaci, G.; and Hafner, V. V. 2016. On the sense of agency and of object permanence in robots. In *2016 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, 166–171. ISSN: 2161-9484.

Bellemare, M. G.; Veness, J.; and Bowling, M. 2012. Investigating contingency awareness using Atari 2600 games. In *Proceedings of the National Conference on Artificial Intelligence*, volume 2, 864–871. ISBN 9781577355687.

Bengio, E.; Thomas, V.; Pineau, J.; Precup, D.; and Bengio, Y. 2017. Independently Controllable Features. ArXiv: 1703.07718.

Blakemore, S. J.; and Frith, C. 2005. The role of motor contagion in the prediction of action. In *Neuropsychologia*, volume 43, 260–267. Elsevier Ltd.

Blakemore, S. J.; Wolpert, D.; and Frith, C. 2000. Why can't you tickle yourself? *Neuroreport*, 11(11): R11–16.

Buesing, L.; Weber, T.; Zwols, Y.; Racaniere, S.; Guez, A.; Lespiau, J.-B.; and Heess, N. 2018. Woulda, Coulda, Shoulda: Counterfactually-Guided Policy Search. ArXiv:1811.06272 [cs, stat].

Choi, J.; Guo, Y.; Moczulski, M.; Oh, J.; Wu, N.; Norouzi, M.; and Lee, H. 2019. Contingency-aware exploration in reinforcement learning. In *7th International Conference on Learning Representations, ICLR 2019*. International Conference on Learning Representations, ICLR.

Corcoll, O.; and Vicente, R. 2020. Disentangling causal effects for hierarchical reinforcement learning. arXiv:2010.01351.

Cullen, K. E. 2004. Sensory signals during active versus passive movement. *Current Opinion in Neurobiology*, 14(6): 698–706.

Fukutomi, M.; and Carlson, B. A. 2020. A History of Corollary Discharge: Contributions of Mormyrid Weakly Electric Fish. *Frontiers in Integrative Neuroscience*, 14.

Gallese, V.; and Goldman, A. 1998. Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, 2(12): 493–501.

Geirhos, R.; Jacobsen, J.-H.; Michaelis, C.; Zemel, R.; Brendel, W.; Bethge, M.; and Wichmann, F. A. 2020. Shortcut learning in deep neural networks. *Nature Machine Intelligence*, 2(11): 665–673. Number: 11 Publisher: Nature Publishing Group.

Haggard, P. 2017. Sense of agency in the human brain. *Nature Reviews Neuroscience*, 18(4): 196–207. Number: 4 Publisher: Nature Publishing Group.

Jékely, G.; Godfrey-Smith, P.; and Keijzer, F. 2021. Reafference and the origin of the self in early nervous system evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1821).

Judea Pearl. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge University Press. ISBN 0-521-77362-8 978-0-521-77362-1.

Kawato, M. 1999. Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, 9: 718–727.

Lapuschkin, S.; Wäldchen, S.; Binder, A.; Montavon, G.; Samek, W.; and Müller, K. R. 2019. Unmasking Clever Hans predictors and assessing what machines really learn. *Nature Communications*, 10(1).

Lewis, M. 2012. *Social cognition and the acquisition of self*. Springer Science & Business Media.

Lin, L.-J. 1992. Self-improving reactive agents based on reinforcement learning, planning and teaching. *Machine Learning*, 8(3): 293–321.

McCarthy, J. 1963. Situations, Actions, and Causal Laws. Technical report, Stanford Artificial Intelligence Project Memo No. 2. DOI 10.21236/AD0785031.

Medendorp, W. P. 2011. Spatial constancy mechanisms in motor control. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1564): 476–491.

Mesnard, T.; Weber, T.; Viola, F.; Thakoor, S.; Saade, A.; Harutyunyan, A.; Dabney, W.; Stepleton, T.; Heess, N.; Guez, A.; Moulines, .; Hutter, M.; Buesing, L.; and Munos, R. 2021. Counterfactual Credit Assignment in Model-Free Reinforcement Learning. ArXiv:2011.09464 [cs].

Miall, R. C.; and Wolpert, D. M. 1996. Forward Models for Physiological Motor Control. *Neural Networks*, 9(8): 1265–1279.

Mnih, V.; Badia, A. P.; Mirza, M.; Graves, A.; Lillicrap, T.; Harley, T.; Silver, D.; and Kavukcuoglu, K. 2016. Asynchronous methods for deep reinforcement learning. In *International conference on machine learning*, 1928–1937. PMLR.

Oh, J.; Guo, X.; Lee, H.; Lewis, R.; and Singh, S. 2015. Action-conditional video prediction using deep networks in Atari games. In *Advances in Neural Information Processing Systems*, volume 2015-Janua, 2863–2871.

Rajmohan, V.; and Mohandas, E. 2007. Mirror neuron system. *Indian Journal of Psychiatry*, 49(1): 66–69.

Roy, J. E.; and Cullen, K. E. 2004. Dissociating Self-Generated from Passively Applied Head Motion: Neural Mechanisms in the Vestibular Nuclei. *The Journal of Neuroscience*, 24(9): 2102–2111.

Sawada, Y. 2018. Disentangling Controllable and Uncontrollable Factors of Variation by Interacting with the World. ArXiv:1804.06955.

Schaul, T.; Quan, J.; Antonoglou, I.; and Silver, D. 2016. Prioritized Experience Replay. ArXiv:1511.05952 [cs].

Schillaci, G.; Ritter, C.-N.; Hafner, V. V.; and Lara, B. 2016. Body Representations for Robot Ego-Noise Modelling and Prediction. Towards the Development of a Sense of Agency in Artificial Agents. 390–397. MIT Press.

Schrittwieser, J.; Antonoglou, I.; Hubert, T.; Simonyan, K.; Sifre, L.; Schmitt, S.; Guez, A.; Lockhart, E.; Hassabis, D.; Graepel, T.; Lillicrap, T.; and Silver, D. 2020. Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model. *Nature*, 588(7839): 604–609. ArXiv:1911.08265 [cs, stat].

Schroder-Schetelig, J.; Manoonpong, P.; and Worgotter, F. 2010. Using efference copy and a forward internal model for adaptive biped walking. 29: 357–366.

Sperry, R. W. 1950. Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Comparative and Physiological Psychology*, 43(6): 482–489.

Straka, H.; Simmers, J.; and Chagnaud, B. P. 2018. A New Perspective on Predictive Motor Signaling. *Current Biology*, 28(5): R232–R243.

Thomas, V.; Bengio, E.; Fedus, W.; Pondard, J.; Beaudoin, P.; Larochelle, H.; Pineau, J.; Precup, D.; and Bengio, Y. 2018. Disentangling the independently controllable factors of variation by interacting with the world. ArXiv:1802.09484.

von Helmholtz, H.; and Southall, J. P. 1924. *Helmholtz's treatise on physiological optics*, volume 1. Optical Society of America.

von Holst, E. 1954. Relations between the central Nervous System and the peripheral organs. *The British Journal of Animal Behaviour*, 2(3): 89–94.

Watson, J. S. 1966. The Development and Generalization of Contingency Awareness in Early Infancy: Some Hypotheses. *Merrill-Palmer Quarterly of Behavior and Development*, 12(2): 123–135.

Wilke, C.; Synofzik, M.; and Lindner, A. 2013. Sensorimotor Recalibration Depends on Attribution of Sensory Prediction Errors to Internal Causes. *PloS one*, 8: e54925.

Wilkins, B.; and Stathis, K. 2022. World of Bugs: A Platform for Automated Bug Detection in 3D Video Games. In *2022 IEEE Conference on Games (CoG)*, 520–523.

Wolpert, D. M.; and Flanagan, J. R. 2001. Motor prediction. In *Current biology*, volume 11, R729–32.

Wolpert, D. M.; Ghahramani, Z.; and Jordan, M. I. 1995. An internal model for sensorimotor integration. *Science*, 269(5232): 1880–1882.

Yang, J.; Lee, G.; Chang, S.; and Kwak, N. 2019. Towards Governing Agent's Efficacy: Action-Conditional β-VAE for Deep Transparent Reinforcement Learning. In *Proceedings of Machine Learning Research*, volume 101, 32–47.

Zaadnoordijk, L.; Besold, T. R.; and Hunnius, S. 2019. A match does not make a sense: on the sufficiency of the comparator model for explaining the sense of agency. *Neuroscience of Consciousness*, 2019(1): niz006.

Zhong, Y.; Schwing, A.; and Peng, J. 2020. Disentangling Controllable Object Through Video Prediction Improves Visual Reinforcement Learning. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, volume 2020-May, 3672–3676. ISBN 9781509066315.