Lecture 08 : Philosophical Issues in Behavioural Science

Stephen A. Butterfill
< s.butterfill@warwick.ac.uk >

Monday, 27th November 2023

Contents

1. Introduction

This lecture introduces two (of many) interface problems. These are problems which arise when actions are controlled by two or more representations that are not inferentially integrated. How is it possible that the two representations non-accidentally match?

Interface problems arise when one action (or event) is controlled by two or more representations that are not inferentially integrated.

The representations' influence on a single action indicates that the outcomes they represent must at least sometimes non-accidentally match. (Unless there is to be nothing at all to coordinate the representations' influence.)

The lack of inferential integration rules out the most straightforward way of explaining how non-accidental matches occur—namely, through processes of inference.

We must therefore ask, How are non-accidental matches possible? The question is an interface problem.

In this lecture we will identify two interface problems. These involve: > * motor representations and intentions (see *The Interface Problem: Motor Representation vs Intention* (section §2)) > * primary motivational states and preferences (see *Preference vs Aversion: A Dissociation* (section §3))

In addition to being a philosophical topic in their own right,[1] interface problems are important for both philosophical and psychological theories of action.

1.1. Prerequisites and What to Skip

This lecture depends on you having studied some sections from a previous lecture:

- ** **ERRoR! MISSING xref FOR** $unit : philosophical_{t}heories_{h}abits$ **
- *Goal-Directed and Habitual Processes* in Lecture 01
- *Motor Representation* in Lecture 07

None of this lecture is required for the minimum course of study.

[1] See Mylopoulos & Pacherie (2016); Fridland (2016); Shepherd (2019); Christensen (2021); or Ferretti & Zipoli Caiani (2021).

2. The Interface Problem: Motor Representation vs Intention

For a single action, which outcomes it is directed to may be multiply determined by an intention and, seemingly independently, by a motor representation. Unless such intentions and motor representations are to pull an agent in incompatible directions, which would typically impair action execution, there are requirements concerning how the outcomes they represent must be related to each other. This is the interface problem: explain how any such requirements could be non-accidentally met.

This part of the lecture was given by Johan Heemskerk: > * slides [pdf] > > * handout [pdf]

2.1. Recap

We have seen arguments for three claims about motor representation:

> Some motor representations represent outcomes rather than, say, only joint displacements and bodily configurations (see *Motor Representation* in Lecture 07).
>
> There are actions whose directedness to an outcome is grounded in motor representation (see *Motor Representations Ground the Directedness of Actions to Goals* in Lecture 07).
>
> Motor representation differs from intention with respect to representational format (see *Motor Representations Aren't Intentions* in Lecture 07).

A consequence of these claims is that a single instrumental action may involve representations of the outcomes to which it is directed in at least two different representational formats, motor and propositional. This leads to what we will call *the interface problem*, which this section introduces.

2.2. The Interface Problem

Realising it is rapidly going cold, you form an intention to drink the tea. Your hand expertly secures the mug and moves it to your mouth exactly as it opens. Nothing is spilled in these exquisitely coordinated movements.

As this illustrates, there are cases in which a particular action is guided both by one or more intentions and by one or more motor representations. In at least some such cases, the outcomes specified by the intentions match the

outcomes specified by the motor representations. Furthermore, this match is not always accidental.

How do non-accidental matches between intention and motor representation come about? (This question is The Interface Problem)

This question is a problem because two natural routes to answering the question are unavailable. Appealing to common causes of intentions and motor representations is a non-starter; and appealing to content-respecting causal processes despite a lack of inferential integration between intentions and motor representations amounts to no more than a stab in the dark.

2.3. Background: Anarchic Hand Syndrome

Marchetti & Della Sala (1998, p. 196) characterise this syndrome as involving:

> 'the occurrence of complex movements of an upper limb which are clearly goal-directed and well executed but unintended (Della Sala et al., 1994). These unwanted movements cannot be voluntarily interrupted and might interfere with the desired actions carried out by the other (healthy) hand. The patients are aware of the bizarre and potentially hazardous behaviour of their hand but cannot inhibit it. They often refer to the feeling that one of their hands behaves as if it has a will of its own, but never deny that this capricious hand is part of their own body. The bewilderment comes from the surprising and unwanted actions, not from a sensation of lack of belonging of the hand.'

For further details, see Young (2013, p. 58ff) (who also quote the above).

2.4. Background: Action Slips

action slips are actions that run contrary to intentions (Norman 1981). For instance:

> 'I was at the end of a salad bar line, sprinkling raisins on my heaping salad, and reached into my left pocket to get a five-dollar bill. The raisins knocked a couple of croutons from the salad to the tray. I reached and picked them up, intending to pop them into my mouth. My hands came up with their respective loads simultaneously, and I rested the hand with the croutons on the tray and put the bill in my mouth, actually tasting it before I stopped myself.' (Norman 1981, p. 10)

For a philosophers' perspective on action slips, see Mylopoulos (2022) (who also introduces many excellent scientific sources).

2.5. Further Reading

Recent work on the Interface Problem about intention and motor representation includes:

- Mylopoulos & Pacherie (2016);
- Fridland (2016);
- Shepherd (2019);
- Christensen (2021);
- Ferretti & Caiani (2019); and
- Ferretti & Zipoli Caiani (2021).

3. Preference vs Aversion: A Dissociation

Your preferences can be incompatible with your aversions (and thereby with primary motivational states). This shows that there is not a single system of preferences in rats or humans.

There are at least two kinds of motivational state, which have distinct roles in explaining behaviour.

One is desires or preferences. These are states that can be influenced by fashion and all kinds of learning.

The other is primary motivational states. These include hunger, thirst, satiety, aversion and disgust. They are linked to biological needs and not always learned. Although they can be modified by learning, there are limits on the influence of learning.

How are primary motivational states and desires related?

Consider two views:

> *Wrong View* Primary motivational states and preferences form a single system for guiding action selection. Unless something goes badly wrong, your primary motivational states inform your preferences directly. To illustrate, if you are hungry for a food, you desire it; and if you are averse to a food, you do not desire it.

In this section we consider evidence against the Wrong View and in favour of:

> *Correct View* There are at least two kinds of motivational state which have distinct roles in explaining behaviour. If primary motivational states inform your preferences at all, they do so

> only indirectly. You may hunger for a food you do not desire to eat, and you may desire to eat a food to which you are averse.[2]

3.1. Anecdote: Dickinson's Water Melons

Dickinson ate melons for the first time. Shortly after consumed red wine and suffered mild toxicosis. Waking up the next day, Dickinson felt thirsty.

> 'once again in need of refreshment, he readily retraced the route to the water melon stall. But when confronted with the direct signal for approach, the sight of those sliced, juicy, rosy-red segments which had looked and were so refreshing on the previous occasion, his appetite abated. And indeed, when he managed to take a bite, he discovered that he had an aversion to water melon with the consequence that none has passed his lips since that day.' (Dickinson & Balleine 1993, p. 285)

The aversion makes sense because becoming unwell after eating a novel food can cause aversion to it (Domjan 2010, p. 71).

Puzzle: if Dickinson was not already averse to melon when he work up, how could eating melon cause aversion? If Dickinson was already averse, why did he set out to eat melon?

Does the Correct View (see above) suggest a solution to this puzzle?

3.2. The Questions

Can your primary motivational states diverge from your preferences?

In particular:

1. Can hunger drive you to seek a novel food even tho you have no desire to eat it? And can satiety reign in your search for a food even though you desire to eat it?
   Yes (Balleine 1992)!
2. And can sugar solution rank highly among your preferences even after you have become averse to it?
   Yes (Balleine & Dickinson 1991)!

[2] Compare Dickinson & Balleine (1995, p. 164): 'a shift in primary motivational state can have little or no direct impact on instrumental performance and thus, by implication, on the incentive value of the outcome. Rather, animals have to learn through experience with a particular food in the undeprived state that it has a low incentive value when they are not hungry.'

The evidence for positive answers to these questions supports the Correct View (see above): Preferences and primary motivational states have distinct roles in explaining behaviour. If primary motivational states inform your preferences at all, they do so only indirectly.

Conclusion: your behaviours are driven by two (or more) motivational systems which are, to an interesting degree, independent of each other.

This conclusion gives rise to an interface problem …

3.3. An Interface Problem

We have seen evidence for these claims:

1. Primary motivational states guide some actions.
2. Preferences guide some actions.
3. Pursuing a single goal can involve both kinds of state, as, for instance, when the rat obtains food by pressing the lever and entering the magazine to retrieve it.
4. Primary motivational states can differ from preferences, as, for instance, when the rat is hungry for the food but has not encountered the food in a hungry state.

The above claims collectively confront us with a challenge. How are non-accidental matchs between preferences and primary motivational states possible?

This question is an interface problem.[3]

This interface problem was raised by Dickinson and Balleine:

> 'we should search in vain among the literature for a consensus about the psychological processes by which primary motivational states, such as hunger and thirst, regulate simple goal-directed [i.e. instrumental] acts' (Dickinson & Balleine 1994, p. 1)

3.4. Background: Outcome-Driven vs Stimulus-Driven

Distinguish two kinds of action-guiding processes:

1. *Outcome-driven processes* are processes guided by expectations concerning how likely an action is to bring about an outcome. They typically depend on action—outcome

[3] We encountered a different interface problem involving motor representations and intentions in *The Interface Problem: Motor Representation vs Intention* (section §2).

links, for example, lever-press—obtain-food. One example is goal-directed processs.

2. *Stimulus-driven processes* are guided by the presence or absence of a stimlus. They typically depend on stimulus—stimulus and stimulus—action links, for example, magazine—food-smell and food-smell—eat. Examples include reflexes and habitual processs.

Warning. I could not find exactly this distinction in any of the textbooks or key sources I checked. Balleine (1992, p. 248) mentions 'Pavlovian processes' but does not say what they are. Domjan (2010, p. 209) comments:

> 'Motivational processes in instrumental behavior have been addressed from two radically different perspectives and intellectual traditions, the associationist perspective rooted in Thorndike's law of effect and Pavlovian conditioning, and the response-allocation perspective rooted in Skinner's behavioral analysis. These two approaches differ in more ways than they are similar, making it difficult to imagine how they might be integrated. The fundamental concept in the associationist approach (the concept of an association) is entirely ignored in the response-allocation approach. Also, the mechanism of response allocation characterized by behavioral economics has no corresponding structure in the associationist approach. Both approaches have contributed significantly to our understanding of the motivation of instrumental behavior. Therefore, neither approach can be ignored in favor of the other.' (Domjan 2010, p. 209)

My proposal, which I take to be implicit in Dickinson and Balleine's work, is twofold. First, outcome-driven and stimulus-driven processes are largely independent of each other. Second, primary motivational states directly influence only stimulus-driven processes.

3.5. Background: Classical vs Operant Conditioning

In operant conditioning, subjects are rewarded when they perform an action in response to a stimulus. (For example, when they press a lever in response to being in a particular cage.) Through operant conditioning, the subject may acquire a habit (stimulus—action link) or learn about the outcomes of actions (action—outcome link).[4]

[4] See Bouton (2016, pp. 28–9) for a simple guide (perhaps too simple).

We encountered operant conditioning in discussing evidence for the existence and influence of habitual processs (see *Goal-Directed and Habitual Processes* in Lecture 01).

In classical conditioning, subjects are exposed to contingencies between stimuli. In Pavlov's famous experiment, a bell was rung each time food appeared. Through classical conditining the subject learns about contingencies between stimuli (stimulus—stimulus link, e.g. bell—food).

On the importance of classical conditioning in human life:

> Classical conditioning is used to study the 'associative learning mechanisms [which] have been shaped by evolution to enable animals to store information about real causal relationships in their environment.' (Dickinson 1980, p. 26)

> 'Classical conditioning is the process whereby we learn to predict when and what we might eat, when we are likely to face danger, and when we are likely to be safe. It is also integrally involved in the learning of new emotional reactions (e.g., fear or pleasure) to stimuli that have become associated with a significant event.' (Domjan 2010, p. 60)

3.6. Significance

Why does the independence of preferences from primary motivational states matter?

In *What Are Preferences?* in Lecture 03 we considered three attempts to anchor the notion of preference or desire. None of those attempts can straightforwardly accommodate the idea that there are two (or more) systems of motivational state.

In *Dual Process Theory Opposes Decision Theory?* in Lecture 03 we considered how to make the dual-process theory of instrumental action consistent with Jeffrey's idea that decision theory can anchor a shared understanding of preference (and subjective probability). One possibility was to think of decision theory as characterising goal-directed processs only. But, as we noted back then, this would require that the rewards needed to characterise habitual processs are not rewards in virtue of your preferences. This is worth reconsidering if there are in fact tow kinds of motivational state which have distinct roles in explaining behaviour.

3.7. Background on Aversion

> 'Pavlovian conditioning can lead to the learning of food preferences and aversions. A taste preference is learned if a flavor is paired with nutritional repletion or other positive consequences (e.g., Capaldi et al. 1997). In contrast, a conditioned taste aversion is learned if ingestion of a novel flavor is followed by an aversive consequence such as indigestion or food poisoning. [...]
>
> 'A growing body of evidence indicates that many human taste aversions are also the result of Pavlovian conditioning (Scalera 2002).[5] [...] The typical aversion learning experience involves eating a distinctively flavored food and then getting sick. [...]
>
> 'food aversion learning can be independent of rational thought processes and can go against a person's own conclusions about the causes of the illness.' (Domjan 2010, p. 71)

3.8. Fun Fact

> 'Classical conditioning was also independently discovered by Edwin Twitmyer in a Ph.D. dissertation submitted to the University of Pennsylvania in 1902 [...] Twitmyer repeatedly tested the knee-jerk reflex of college students by sounding a bell 0.5 seconds before hitting the patellar tendon just below the knee cap. After several trials of this sort, the bell was sufficient to elicit the knee-jerk reflex in some of the students.' (Domjan 2010, p. 60)

4. Interface Problems and the Role of Experience

Is there a role for experience in solving interface problems? On the interface problem involving preferences and primary motivational states, Dickinson & Balleine (1994) suggest a solution involving experience of bodily reactions to stimuli. And perhaps their idea could be extended and adapted to solve the interface problem involving motor representations and intentions.

We have now encountered two interface problems:

- one involving motor representations and intentions (see *The Interface Problem: Motor Representation vs Intention* (section §2)).

[5] Actually the source cited stresses that there are few human studies. 'The CTA [conditioned taste aversion] has been extensively investigated in a wide variety of laboratory and wild animal species but only incidentally in humans' (Scalera 2002, p. 168).

- one involving preferences and primary motivational states (see *Preference vs Aversion: A Dissociation* (section §3)); and

Is there a role for experience in solving the latter interface problem?

Consider this proposal:

1. Aversion, hunger and other primary motivational states modulate bodily responses to stimuli.
2. Those bodily responses can be, and often are, experienced. For instance, encountering a food you are averse to might trigger peculiar feelings.
3. These expeirences have valence.
   For example, bodily responses caused by aversion are typically experienced as unpleasant.
4. The valence of the feelings influences your preferences.
   For example, however much you might want to eat a food initially, you will probably want to eat it a lot less if aversion to it reliably triggers bodily responses experienced as unpleasant.

This seems to be roughly what Dickinson and Balleine are suggesting:

> 'primary motivational states, such as hunger, do not determine the value of an instrumental goal directly; rather, animals have to learn about the value of a commodity in a particular motivational state through direct experience with it in that state.' (Dickinson & Balleine 1994, p. 7)

And:

> 'the assignment of incentive value is based on learning about one's own hedonic or affective reactions to the goal, reactions that are modulated by primary motivational states.' (Dickinson & Balleine 1995, p. 166)

4.1. Objection

On this proposal, cognition would be inefficient since it relies on experience to sync your preferences with your primary motivational states. But animal cognition is not typically inefficient. So this proposal is probably wrong.

4.2. Reply: Loose Coupling

A basic advantage of any dual-process theory is that it permits loose coupling.

In the case of preferences and primary motivational states, loose coupling is useful because

- your primary motivational states keep you mostly on track
  by, for example, preventing you from starving; and
- loose coupling means that you can draw on your learning
  to pursue things which appear harmful but are actually beneficial (such as chemotherapy)
  and to avoid things which appear beneficial but are actually harmful (such as opiates).

Compare Dickinson and Balleine:

> 'the motivational control over goal-directed actions is, at least in part, indirect and mediated by learning about one's own reactions to primary incentives. By this process [...], goal-directed actions are liberated from the tyranny of primary motivation.' (Dickinson & Balleine 1994, p. 16)

4.3. Speculative Extension

These ideas might motivate considering whether there is a role for experience in solving the other interface problem about motor representations and intentions (see *The Interface Problem: Motor Representation vs Intention* (section §2)):

1. Motor representations of outcomes structure experiences, imaginings and (prospective) memories
   in ways which provide opportunities for attention to actions directed to those outcomes.
2. Forming intentions concerning an outcome can influence attention to the action,
   which can influence the persistence of a motor representation of the outcome.

5. Conclusion

One action can involve multiple, dissociable motivational and effective states.

We do not understand how these ever nonaccidentally match, although experience of our own bodies and of action possibilities may play a role.

Glossary

action slip 'A slip is a form of human error defined to be the performance of an action that was not what was intended' (Norman 1981, p. 1). Examples include saying *canpakes* for *pancakes* or pouring coffee on to cereal. 4

anchor A theory, fact or other thing that is used by a group of researchers to ensure that they have a shared understanding of a phenomenon. An anchor is needed when it is unclear whether different researchers are offering incompatible claims about a single phenomenon or compatible claims about distinct phenomena. For example, we might take decision theory to anchor a shared understanding of belief and desire. 9

classical conditioning Also called 'Pavlovian conditioning'. A stimulus—stimulus link is learned through explosure to contingencies, causing you to respond to the first stimulus as if the second were present. Pavlov famously conditioned a dog to salivate on hearing a bell by having the bell sound before food arrived. Subject to effects like blocking and overshadowing. 9, 15

decision theory I use 'decision theory' for the theory elaborated by Jeffrey (1983). Variants are variously called 'expected utility theory' (Hargreaves-Heap & Varoufakis 2004), 'revealed preference theory' (Sen 1973) and 'the theory of rational choice' (Sugden 1991). As the differences between variants are not important for our purposes, the term can be used for any of core formal parts of the standard approaches based on Ramsey (1931) and Savage (1972). 9, 13

dual-process theory Any theory concerning abilities in a particular domain on which those abilities involve two or more processes which are distinct in this sense: the conditions which influence whether one mindreading process occurs differ from the conditions which influence whether another occurs. 12

dual-process theory of instrumental action Instrumental action 'is controlled by two dissociable processes: a goal-directed and an habitual process' (Dickinson 2016, p. 177). (See instrumental action.) 9

goal-directed process A process which involves 'a representation of the causal relationship between the action and outcome and a representation of the current incentive value, or utility, of the outcome' and which influences an action 'in a way that rationalizes the action as instrumental for attaining the goal' (Dickinson 2016, p. 177). 8, 9

habitual process A process underpinning some instrumental actions which obeys *Thorndyke's Law of Effect*: 'The presentation of an effective [=rewarding] outcome following an action [...] reinforces a connection between the stimuli present when the action is performed and the action itself so that subsequent presentations of these stimuli elicit the [...] action as a response' (Dickinson 1994, p.48). (Interesting complication which you can safely ignore: there is probably much more to say about under what conditions the stimulus–action connection is strengthened; e.g. Thrailkill et al. 2018.) 8, 9

inferential integration For states to be *inferentially integrated* means that: (a) they can come to be nonaccidentally related in ways that are approximately rational thanks to processes of inference and practical reasoning; and (b) in the absence of obstacles such as time pressure, distraction, motivations to be irrational, self-deception or exhaustion, approximately rational harmony will characteristically emerge, eventually, among those states. 2, 4

instrumental action An action is *instrumental* if it happens in order to bring about an outcome, as when you press a lever in order to obtain food. (In this case, obtaining food is the outcome, lever pressing is the action, and the action is instrumental because it occurs in order to bring it about that you obtain food.) You may encounter variations on this definition of *instrumental* in the literature. For instance, Dickinson (2016, p. 177) characterises instrumental actions differently: in place of the teleological 'in order to bring about an outcome', he stipulates that an instrumental action is one that is 'controlled by the contingency between' the action and an outcome. And de Wit & Dickinson (2009, p. 464) stipulate that 'instrumental actions are *learned*'. 3, 13

interface problem An interface problem may arise when two kinds of representation sometimes non-accidentally match: the problem is to explain how such matches are possible. 2, 7, 10, 12

match [of outcomes] Two collections of outcomes, A and B, *match* in a particular context just if, in that context, either the occurrence of the A-outcomes would normally constitute or cause, at least partially, the occurrence of the B-outcomes or vice versa.

To illustrate, one way of matching is for the B-outcomes to be the A-outcomes. Another way of matching is for the B-outcomes to stand to the A-outcomes as elements of a more detailed plan stand to those of a less detailed one.

[of plan-like structures] In the simplest case, plan-like hierarchies of motor representations *match* if they are identical. More generally, plan-like hierarchies *match* if the differences between them *do not matter* in the following sense. For a plan-like hierarchy in an agent, let the *self part* be those motor representations concerning the agent's own actions and let the *other part* be the other motor representations. First consider what would happen if, for a particular agent, the other part of her plan-like hierarchy were as nearly identical to the self part (or parts) of the other's plan-like hierarchy (or others' plan-like hierarchies) as psychologically possible. Would the agent's self part be different? If not, let us say that any differences between her plan-like hierarchy and the other's (or others') are *not relevant* for her. Finally, if for some agents' plan-like hierarchies of motor representations the differences between them are not relevant for any of the agents, then let us say that the differences *do not matter*.

[of motivational states] Two motivational states match in a particular context just if, in that context, the actions one would cause and the actions the other would cause are all proper ways of fulfilling both motivational states. 2, 3, 7

motor representation The kind of representation characteristically involved in preparing, performing and monitoring sequences of small-scale actions such as grasping, transporting and placing an object. They represent actual, possible, imagined or observed actions and their effects. 2, 3, 7, 10, 12

operant conditioning Also called 'instrumental conditioning'. A stimulus—action link is learned through your action being rewarded when it occurs with the stimulus. 8, 9

primary motivational state A state such as hunger, thirst, satiety, aversion or sexual arousal. Primary motivational states are closely linked to biological needs. They are not all acquired through learning; and learning has limited effects on them, although classical conditioning can modify them (Capaldi et al. 1997). 2, 5, 7–11

problem a question that is difficult to answer. 4

References

Balleine, B. (1992). Instrumental performance following a shift in primary motivation depends on incentive learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *18*, 236–250.

Balleine, B. & Dickinson, A. (1991). Instrumental performance following reinforcer devaluation depends upon incentive learning. *The Quarterly Journal of Experimental Psychology Section B*, *43*(3), 279–296.

Bouton, M. E. (2016). *Learning and Behavior: A Contemporary Synthesis* (Second Edition, Second Edition ed.). Oxford, New York: Oxford University Press.

Capaldi, E. D., Hunter, M. J., & Lyn, S. A. (1997). Conditioning with taste as the CS in conditioned flavor preference learning. *Animal Learning & Behavior*, *25*(4), 427–436.

Christensen, W. (2021). The Skill of Translating Thought into Action: Framing The Problem. *Review of Philosophy and Psychology*, *12*(3), 547–573.

de Wit, S. & Dickinson, A. (2009). Associative theories of goal-directed behaviour: A case for animal–human translational models. *Psychological Research PRPF*, *73*(4), 463–476.

Dickinson, A. (1980). *Contemporary animal learning theory*. Problems in the behavioural sciences. Cambridge: Cambridge University Press.

Dickinson, A. (1994). Instrumental conditioning. In N. Mackintosh (Ed.), *Animal Learning and Cognition*. London: Academic Press.

Dickinson, A. (2016). Instrumental conditioning revisited: Updating dual-process theory. In J. B. Trobalon & V. D. Chamizo (Eds.), *Associative learning and cognition*, volume 51 (pp. 177–195). Edicions Universitat Barcelona.

Dickinson, A. & Balleine, B. (1993). Actions and responses: the dual psychology of behaviour. In N. Eilan, R. McCarthy, & B. Brewer (Eds.), *Spatial representation: problems in philosophy and psychology* (pp. 277–293). Oxford: Oxford University Press.

Dickinson, A. & Balleine, B. (1994). Motivational control of goal-directed action. *Animal Learning & Behavior*, *22*(1), 1–18.

Dickinson, A. & Balleine, B. (1995). Motivational Control of Instrumental Action. *Current Directions in Psychological Science*, *4*(5), 162–167.

Domjan, M. (2010). *The Principles of Learning and Behavior*. Wadsworth Cenage Learning.

Ferretti, G. & Caiani, S. Z. (2019). Solving the Interface Problem Without Translation: The Same Format Thesis. *Pacific Philosophical Quarterly*, *100*(1), 301–333.

Ferretti, G. & Zipoli Caiani, S. (2021). How Knowing-That and Knowing-How Interface in Action: The Intelligence of Motor Representations. *Erkenntnis*, 1–31.

Fridland, E. (2016). Skill and motor control: Intelligence all the way down. *Philosophical Studies*, *174*(6), 1539–1560.

Hargreaves-Heap, S. & Varoufakis, Y. (2004). *Game theory: a critical introduction*. London: Routledge.

Jeffrey, R. C. (1983). *The Logic of Decision, second edition*. Chicago: University of Chicago Press.

Marchetti, C. & Della Sala, S. (1998). Disentangling the Alien and Anarchic Hand. *Cognitive Neuropsychiatry*, *3*(3), 191–207. _eprint: https://doi.org/10.1080/135468098396143.

Mylopoulos, M. (2022). Oops! I Did it Again: The Psychology of Everyday Action Slips. *Topics in Cognitive Science*, *14*(2), 282–294.

Mylopoulos, M. & Pacherie, E. (2016). Intentions and Motor Representations: The Interface Challenge. *Review of Philosophy and Psychology, forthcoming*, 1–20.

Norman, D. A. (1981). Categorization of action slips. *Psychological Review*, *88*(1), 1–15.

Ramsey, F. (1931). Truth and probability. In R. Braithwaite (Ed.), *The Foundations of Mathematics and Other Logical Essays*. London: Routledge.

Savage, L. J. (1972). *The Foundations of Statistics* (2nd rev. ed ed.). New York: Dover Publications.

Scalera, G. (2002). Effects of Conditioned Food Aversions on Nutritional Behavior in Humans. *Nutritional Neuroscience*, *5*(3), 159–188.

Sen, A. (1973). Behaviour and the Concept of Preference. *Economica*, *40*(159), 241–259.

Shepherd, J. (2019). Skilled Action and the Double Life of Intention. *Philosophy and Phenomenological Research*, *98*(2), 286–305.

Sugden, R. (1991). Rational Choice: A Survey of Contributions from Economics and Philosophy. *The Economic Journal*, *101*(407), 751–785.

Thrailkill, E. A., Trask, S., Vidal, P., Alcalá, J. A., & Bouton, M. E. (2018). Stimulus control of actions and habits: A role for reinforcer predictability and attention in the development of habitual behavior. *Journal of Experimental Psychology: Animal Learning and Cognition*, *44*, 370–384.

Young, G. (2013). *Anarchic Hand*, (pp. 57–73). Palgrave Macmillan, London. [Online; accessed 2023-11-26].