Computer Science & Philosophy

Computer Science & Philosophy

Prof. Peter Millican, Hertford College

Introducing Some Philosophical Topics, and Links Between the Disciplines



1

Agenda

1. Six topics discussed in the *General Philosophy* lectures available from: http://podcasts.ox.ac.uk/series/general-philosophy-2018 (introductory lectures for all Oxford students in Philosophy).
2. How Computing/AI is relevant to Philosophy (including some of these topics), both in subject-matter and by assisting our understanding.
3. Illustrating this with four examples that relate to *free will*, *motivation*, and *morality* (which are also linked to the worksheet provided).
4. Links to other resources.

2

2

Six Topics and their History

- The topics are *mind and body*, *personal identity*, *action and motivation*, *free will*, *the nature of morality*, and *God and evil*.
  - These have been discussed since ancient times, and remain controversial today. They're closely related in many ways, and it can be fascinating to see how they interact in the views of major thinkers.
  - The lectures include discussion of Aristotle, Galileo, Descartes, Hobbes, Locke, Newton, Butler, Hume, Darwin and Einstein, as well as many recent thinkers.



3

3

Topic 1: Mind and Body

- What is the nature of the thinking *mind*, and its relationship to the body?
  - Does thinking involve something quite distinct from the body (e.g. an immaterial spirit or soul), or is it just a process of the material brain?
  - How do bodily and mental events interact (e.g. when I feel a pain on touching a thorn, or when I decide to raise my arm)?
  - What is the nature of consciousness, and which other living things experience it?
  - Could a computer think in much the same way as we do? Could it be conscious?

4

4

Topic 2: Personal Identity

- What is it for a person to persist through time, and what makes me the *same* person as I was yesterday, last year, or when I was a baby?
  - Our bodies and minds are changing all the time, as we breathe, eat, think, feel etc. We can even undergo major surgery or transplants. But how much change is compatible with remaining *the same person*?
  - Suppose "I" wake up one day with your thoughts and memories instead of my own (and *vice-versa*)? Which of those two people would really be *me*?
  - Is Star-Trek-style teleportation a theoretical possibility? Would that mean that people could be duplicated?
  - Could a computer be a "person" in the relevant sense?

5

5

Topic 3: Action and Motivation

- What motivates us to act as we do? Are we all ultimately self-interested, or can we be genuinely *altruistic* (i.e. favouring others over ourselves)?
  - The theory of *psychological egoism* claims that even when we seem to be aiming to do good for others (e.g. family or friends), we really have in mind a benefit to ourselves (e.g. achieving our own pleasure from witnessing their pleasure). Is this theory plausible?
  - Is it specially rational to act for my own benefit (and perhaps irrational to act for the benefit of others)? If so, What is it about *my future self* that makes this the case?
  - How would these considerations play out if my mind switched with yours, or if I were Star-Trek-duplicated?

6

6

Computer Science & Philosophy

Topic 4: Free Will

- We think of ourselves as acting "freely", but is this just an illusion? And if it is, what impact does that have on moral responsibility?
  - Is everything that happens in the world the inevitable causal result of previous events and laws of nature? If so, does that mean that every action I perform could in principle have been predicted, even before I was born?
  - In that case, I'm just a plaything of fate: so how can I properly be held morally responsible for what I do?
  - Maybe some events (e.g. quantum events) are random, but how can that help to confer moral responsibility?
  - Are other animals capable of free action? And could a robot be just as free as we are?

7

7

Topic 5: The Nature of Morality

- What is the basis of morality, where does it come from, and why should we take it seriously?
  - Is morality "objective", in the sense that moral claims are true or false, independently of what anybody thinks about them? If so, how can moral truth be discovered?
  - Or is morality "subjective", e.g. a matter of opinion or perhaps cultural indoctrination? If so, are moral rules genuinely obligatory (even if we disagree with them)?
  - Why, in either case, should we take any notice of morality? Is it in our own self-interest to be moral?
  - Could a robot be either morally assessable for how it behaves, or worthy of moral consideration by us?

8

8

Topic 6: God and Evil

- Is there a God in the traditional monotheistic sense (i.e. an all-powerful, all-knowing, perfectly good creator of the Universe)? And if there is, how can the evil in the world be explained?
  - What reasons are there for believing, or disbelieving, in such a God, and how strong are they?
  - Could the existence of God ground a satisfactory account of morality, as an expression of the divine will?
  - If there is a God, could evil in the world be explained as arising from the free actions of humans (or other agents, e.g. fallen angels), on the basis that (a) God has reason to create such morally significant agents, and (b) if they are responsible for evil happenings, then God is not?

9

9

Philosophy and Computing/AI

- We've already seen numerous questions raised by the prospect of artificial intelligence:
  - Could a computer think in the same way that we do?
  - Could a computer be conscious?
  - Could a computer count as a "person", with something like personal identity over time?
  - Could a robot be capable of free action, as we are?
  - Could a robot legitimately be morally assessable (e.g. praised or blamed) for how it behaves?
  - Could a computer or robot be worthy of moral consideration by us (e.g. so we avoid harming it)?

10

10

- But there are also plenty of other links between Computing and Philosophy, in many different areas:
  - *Logic*: At the core of both disciplines, with numerous issues of interest to both computer scientists and philosophers.
  - *Ethics of AI*: Many important issues such as algorithmic bias, automation and unemployment, data security, deep fakery, privacy, propaganda, robot ethics, safety, superintelligence.
  - *Philosophy of Science*: How far should computer models (e.g. of global warming) be trusted?
  - *Philosophy of Social Science*: Can computer models revolutionise the study of Economics and Politics?
  - *Philosophy of Mathematics*: Should we accept computer-generated proofs we cannot fully grasp?
  - *Aesthetics*: Can a computer create genuine art?
  - *New Philosophical Fields*: Philosophy of Artificial Intelligence, Philosophy of Artificial Life, Philosophy of Computation, Philosophy of Information, ...

11

11

Computer Models: Prostheses for the Mind

- Another quite different way in which Computing is ***uniquely*** valuable to Philosophy is making it possible to model a wide range of phenomena.
- This gives a wonderful new tool for thought-experiments, exploring possibilities that are too complex to calculate "by hand", promoting both discovery and understanding.
- For lots of online examples, see https://www.turtle.ox.ac.uk/csac (or download the *Turtle System*).



12

12

Peter Millican, Hertford College, Oxford, January 2021

Computer Science & Philosophy

Some Issues Around Free Will, Motivation and Responsibility

- These are big topics (filling Lectures 6 and 7 in the online *General Philosophy* series). Here we'll just be looking at a few snapshots to draw attention to some key points on:
  a) Determinism and predictability
  b) The possibility of multiple explanations
  c) Self-interest, evolution and morality
- If you find this material interesting, please see the online lectures! There's also lots of related material at http://www.philocomp.net/videos.htm.

13

13

(a) Determinism and Predictability

- Determinism is the thesis that all events are "determined" by prior causes. So for any event *E*, given the relevant causal laws and the prior state of the universe, *E* was inevitable.
- It's tempting to think that if determinism is true, and if we have worked out what the relevant causal laws are, then any event should be predictable, at least "in principle" (though of course we might not yet have developed the necessary measuring or calculating equipment).
- This, surprisingly, has turned out to be false!

14

14

The Discovery of "Chaos"

- Mathematician Henri Poincaré saw hints of chaos in the 1880s, and the early 20th century brought some studies of phenomena that are "non-linear" (i.e. not governed by linear equations such as $y = ax + b$, which have the property that when combined their results remain linear).
- But it wasn't until the 1960s, through computer modelling, that chaotic *sensitive dependence on initial conditions* came to prominence.
- In 1976, Robert May wrote an influential paper on the (non-linear) *logistic equation*.

15

15

The Logistic Equation

- Suppose some environment supports an absolute maximum of 100,000 insects, and the current population, expressed as a proportion of 100,000, is P (e.g. if the population is 47,200, $P = 0.472$).
- In each generation, the insects reproduce by some multiplier r (e.g. $r = 3.8$), except that the population is also reduced, owing to habitat depletion, by a factor of $(1 - P)$.
- Hence the population in the next generation is given by the so-called *logistic equation*:

$$P' = r\,P\,(1 - P)$$

16

16

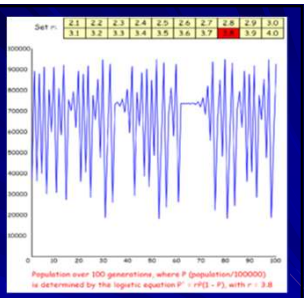
Experimenting with the Logistic Equation

- Go to the "self-similarity and chaos" section of the *Turtle System* "Examples" menu (or if using *Turtle Online*, go to chapter 7 at https://www.turtle.ox.ac.uk/csac)
- Select "Logistic equation" or "Logistic spider" and click on "RUN".
- Then click on different values of *r* (at the top of the canvas) to see the effect.

17

17

Chaos and Quasi-Randomness



A surprising discovery is that if *r* is 3.57 or greater, iterations of the logistic equation generate *chaotic* behaviour, in which the pattern of development exhibits *sensitive dependence on initial conditions*. So even if things are actually entirely deterministic, they look random (and, given the limits of our measurements, they cannot be predicted – or even retrospectively explained except in general terms – on a deterministic basis).

18

18

Peter Millican, Hertford College, Oxford, January 2021

Computer Science & Philosophy

(b) Multiple Explanations

- A vitally important point in considering free action is that a single event or outcome can potentially be explained in more than one way.
- Our first example is taken from game-playing, and illustrates how *purposive explanation* differs from, but is compatible with, *physical causal explanation* (even in a physical system)
- Our second example illustrates *evolutionary explanation* where causal explanation is practically impossible, even though the system may be causally determined.

19

19

Causes and Reasons

- Here's a popular argument against determinism:
  - "You argue that determinism is true. But if you're right, and what you say is causally determined, then the reason you're saying those words is that your brain and body are causing you to say them, not because they're actually true or rationally based. So on your own principles, I've got no reason for taking your argument for determinism seriously."
- This fallaciously treats causal explanation and rational explanation as mutually exclusive. But a chess-playing computer gives a clear illustration of why this is quite wrong.

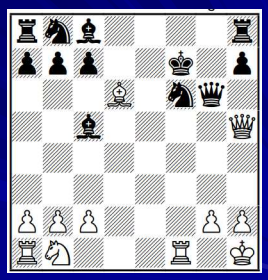
20

20

Purposive Explanation

"Why did the computer play
15. Rxf6+ ?"

"Because then Black can only save the queen with
15. … Kxf6
after which White wins by force with
16. Qe5+ Kf7
17. Qe7+ Kg8
19. Qf8 checkmate."



21

21

- The chess computer is clearly a physical system, operating according to the laws of mechanics and electronics etc., and yet it is set up in such a way that its behaviour conforms to good chessplaying.
  - For an example of how "intelligent" gameplaying can be programmed, see http://www.philocomp.net/videos.htm
- It would clearly be fallacious to argue:
  "The computer chose Rxf6+ because it was physically determined to do that, so it didn't choose it on any *logical ground*, and hence we have no reason for taking its decision seriously."
- Presumably our brains have evolved so that our reasoning likewise generally responds to rational constraints (that's why brains are so useful!).

22

22

Illustrating Evolutionary Explanation

- A long-standing puzzle, unexplained until 1930, is why species such as cattle give birth with roughly equal frequency to males and females. A herd is better off if it consists mainly of cows (peacefully producing and nurturing young etc.) than if it has lots of bulls (who don't help, and are likely to spend much of their time fighting for access to the cows).
  - So why hasn't evolution – which is generally thought to lead to "survival of the fittest" – solved this problem by giving cattle a "fitter" sex ratio?
- Computer modelling can show that the true explanation derives from the logic of evolution.

23

23

A Computerised Thought-Experiment

- Imagine a sexually reproducing species that can influence the sex of its offspring through a "Female Offspring Probability" (FOP) gene:
  - This determines the probability that any new offspring will be female (from 0% to 100%).
  - The gene is inherited in the usual way from one of the parents, and subject to small (quasi-) random variation, say ± 10%, in its probability value.
  - Mating between males and females is (quasi-) random, as is death of individuals.
  - We start the simulation with all FOP values initialised to 90%+, and run it …

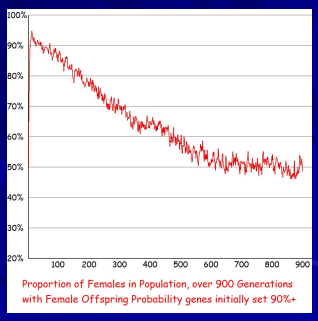
24

24

Computer Science & Philosophy

Running the Sex Ratio Program

Owing to the high FOP values, the proportion of females quickly rises to more than 90%. But then the proportion gradually goes lower and lower, until it gets to around 50% - after which it stays there (with continuous small random variations).



25

25

Why Do The Sexes Equalise?

- (One Level of) Causal Explanation
  - "Female F_{61}, with a FOP gene of 93%, mated with male M_{273}, with a FOP of 97%, producing new female F_{501}, inheriting FOP value of 94%; then female F_{372}, …" [and so on for hundreds of thousands more]
- The Evolutionary Explanation
  - If the population is mostly female, then males will on average have more offspring than females. Hence those individuals who have a low-value FOP gene – and hence have more male offspring – will tend to have more ***grandchildren*** than others, and hence their low-value FOP genes will spread preferentially …

26

26

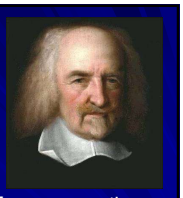
(c) Self-Interest, Evolution and Morality

- The theory of psychological egoism – that we always act with a view to our own self-interest – is very popular with armchair cynics. But fortunately, *it can be shown to be clearly false*: see General Philosophy lecture 6.
  - One obvious objection comes from evolution, since animals that care for their offspring will tend to leave more offspring than those that don't care; hence genes favouring such caring will spread.
  - But this only explains why we should care for our own offspring. It's more difficult to explain how evolution could lead us to care for others more widely.

27

27

Hobbes' *Leviathan* (1651)



- More than two centuries before Darwin, Thomas Hobbes (alumnus of Hertford College!) was trying to solve a related problem: how can a stable society arise from competing selfish individuals, all desperately trying to avoid a "state of nature" – a war of all against all – in which life would be "solitary, poor, nasty, brutish, and short"?
- His simple "thought-experiments" regarding how this might be achieved concluded that the only solution was for everyone to club together and agree to obey *an absolute sovereign* (such as Charles I or Cromwell's Parliament – this was the time of the English Civil War).

28

28

If Only Hobbes Had Been Able to Run Computer Models!

- Computer models can demonstrate better solutions to Hobbes' problem of achieving cooperation. We start from the famous *Prisoner's Dilemma*, in which rational self-interest seems to exclude cooperation:

	You Cooperate	**You Defect**
I Cooperate	We both get £3	I get nothing You get £5
I Defect	I get £5 You get nothing	We both get £1

29

29

The Prisoner's Dilemma and Evolution

- In the Prisoner's Dilemma (PD), the parties act quite separately, and the only self-interestedly "rational" action seems to be to defect.
- It might therefore seem that evolution in this scenario could only favour defection behaviour.
- The PD is a "dilemma" because if both parties follow self-interest in this way, both do worse than if they had "altruistically" cooperated.
- Moreover in nature cooperation does seem to occur. Might there be some natural way in which the benefits of cooperation can be achieved?

30

30

Computer Science & Philosophy

Iterating the Prisoner's Dilemma

- In nature, we encounter each other *repeatedly*, and we follow *patterns* of behaviour rather than calculating every decision separately.
- What happens if we allow the interactions to be *iterated* (i.e. repeated), say 10 times, so that individuals can condition their responses in the light of previous interactions?
- Suppose we also put this in an evolutionary or other competitive context, where behaviours can evolve, or be transmitted (e.g. by imitation or learning) between different individuals?

31

Three Simple IPD Strategies

- ALLC – naïve cooperator: *always cooperates*
- ALLD – nasty defector: *always defects*
- TFT – tit-for-tat: *cooperates* on the first iteration, *then reciprocates* whatever the other player did on the previous iteration.
- Over 10 iterations, pairwise totals will be:

ALLD v. ALLC:	£50 – £0
ALLD v. ALLD:	£10 – £10
ALLD v. TFT:	£14 – £9
Other pairings (ALLC/TFT):	£30 – £30

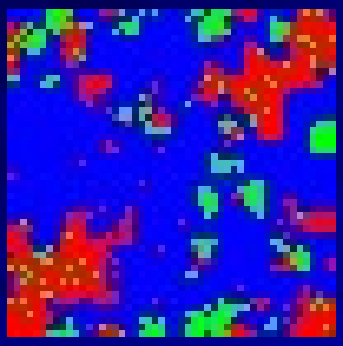
32

Running the IPD Model

- We start with a random 2% of individuals being nasty defectors (blue), 14% tit-for-tats (green), and the rest naïve cooperators (red).
- Each individual interacts for 10 iterations with each of its 8 neighbours (with the grid "wrapping around" from edge to edge), keeping track of its total score from all 8 of these series.
- When all total scores are known, each cell gets "taken over" by whichever strategy achieved the largest total score in its 9-cell neighbourhood.
- Repeat … and see what happens!

33

Snapshot of an IPD



Every cell has a sequence of 10 interactions with each of its 8 neighbours. Then the most successful strategy in the neighbourhood "wins the cell". Most were initially naïve cooperators (RED), but the nasty defectors (BLUE) have ruthlessly taken advantage of them. Small groups of tit-for-tats (GREEN) survive, because they support each other and don't allow the defectors to exploit them beyond the first turn. Before long, the greens will dominate and the blues disappear.

34

34

Drawing Morals

- Let's draw some tentative conclusions:
  - Causal determinism doesn't imply predictability, and nor does it conflict in any way with behaviour that is purposive and/or rational. Perhaps it's less relevant to "free will" than commonly thought?
  - Evolution can explain not only how our bodies work, but also how we think: giving us desires and feelings that concern others as well as ourselves.
  - These could include "moral" feelings (e.g. of guilt, blame, or gratitude), adapting us to fit into human society, and perhaps so "hard-wired" into us that they are unaffected by theoretical considerations.

35

35

Related Resources

- www.philocomp.net Lots of stuff on Computer Science & Philosophy, including more recorded video talks.
- www.turtle.ox.ac.uk Home of *The Turtle System*, *Turtle Online*, and *Computer Science Across the Curriculum* (CSAC). Free software and learning materials for programming, with loads of illustrative programs taken from across the school curriculum (biology, chemistry etc.), and plenty that have philosophical interest.
- http://podcasts.ox.ac.uk/series/general-philosophy-2018 Eight General Philosophy lectures covering the first year Oxford course, and presupposing no prior knowledge of Philosophy. (The first year CS&P lectures on Alan Turing are also available, at series/alan-turing-2018.)

36

36

Peter Millican, Hertford College, Oxford, January 2021