KCL Advanced Research Seminar

Two diachronic grounds for movement within conceptual spaces

Nathan Oseroff
King's College London
nathan.oseroff@kcl.ac.uk

A connection between pedagogical and epistemic problems

- I will give you two explanations for theory-preference that rely on diachronic justifications for
  1. the simplest available theory
  2. that also preserves the greatest content of our previous theoretical web and
  3. takes account of the available accepted evidence.

- One explanation will be naturalistic or psychologistic (Gärdenfors 2014) and is aimed at teaching (guided inquiry).
- The other explanation will be computational and long-run truth-directed (Kelly 2004) and is aimed at unguided inquiry (current work in the sciences).

- What unites the naturalistic and computational approaches are rules for movement that
  - minimise a long-term issue: the number of theories (and corresponding conceptual spaces) we transition between as we engage in inquiry
  - while maintaining other short-term goals: satisfying simplicity and continuity of content.

- Starting with teaching is a helpful toy example:
    - Teachers deal with *known* starting and endpoints, and the pedagogical rules set constraints within *previously explored* concept-space.
- It's far more difficult once we turn towards the sciences:
    - Scientists deal with an unknown endpoint and attempt to follow epistemic rules to constrain the possible set of *future moves* in concept-space.

- Rules on inquiry regulate the movement between *conceptual spaces* and the overall *efficiency of the path*.
- This can be modelled using Gärdenfors' work (2000) and directional graphs, representing the past development of a scientific research programme:
    - it provides naturalistic grounds for which step must be taken, i.e. the mental *ease* of acquisition of new concepts.

- Gärdenfors' work relies on producing a measurement of (e.g. in Euclidean space) as a function of *distance* between any two points.
    - If B is closer than C to A, then B is more similar to A than C.
- With this rule, Gärdenfors produces a geometric model of simplicity: we can construct a Voronoi tessellation of the space, which is a process that breaks the space up into convex regions with the aid of *prototypes*.

- For most models constructed in Euclidean space, these will be representations of our mental categories that are simplest, given any number of prototypes.
- So long as a Voroni tessellation is maintained, any introduction of a new prototype in this space
  - will bear the greatest similarity to the previous conceptual space, given the new prototype.
  - will be the simplest available tessellation, given the new prototype.

- In contrast to Zenker and Gärdenfors (2014), I see a *realist* interpretation, rather than instrumentalist:
    - there is a correspondence relation between scientific theories (T), conceptual spaces (CS) and the world
    - there is a correspondence relation between T and CS
    - T and CS may be more or less empirically adequate
    - T and CS may be more or less simple
    - T and CS may be more or less similar to other T and CC

The pedagogical problem and Gärdenfors

- Consider this question: How should teachers better introduce new concepts to students in a way that provides the *least cognitive stress*? We want
  - the *simplest* CS at each stage
  - the *fewest* CS possible
  - adoption of the *most similar* CS to our previous CS

How not to solve the problem

- Consider the following distinction between synchronic and diachronic perspectives:
  - A synchronic theory describes relations of support and coherence between a system (of beliefs, theories, concepts) at a single time
  - A diachronic theory describes changes (to beliefs, theories) over time
- It's reasonable to have both kinds of theory at our disposal, but we want a helpful balance of the two and not neglect one at the expense of the other.

- If there is too much emphasis on a synchronic perspective these attempts start at the *end product of previous inquiry*: we guide inquiry by attempting to maximise true beliefs, maintain coherence, and minimise false beliefs by laying out what our currently best models are.
- Many of the explanations for why we value particular epistemic norms rely on the synchronic side at the expense of the diachronic side.

- If inquiry were entirely synchronic-oriented, we would want to maximise true beliefs and limit exposure to false beliefs.
    - There would be little talk of our past mistakes.
- Lastly, we want to retain coherence.
    - But much of history of science is about the discovery of incoherence between theory and the world.

It doesn't reflect teaching

- Our very models are known to be false: they are often abstractions that provide conceptual ease to their use.
- Teachers work with these historical fictions because they ease students from one conceptual space to another.
- Much of the learning experience is coming to grips with the failure of coherence between theory and reality.

- What have we learned by examining an obviously absurd scenario?
  - Teachers cannot intelligibly communicate to students using concepts that differ *too* much from whatever conceptual spaces they presently use.
  - In order to arrive at that end state, we cannot do so in one step, but through intermediary steps that maintain similarity between each conceptual space.
- *How many steps maximises our three goals?*

- An analogy: although many routes lead to Rome, the best route for us to take at any one time may be unique.
- What is the most appropriate route for students to take from their starting point? How should teachers help guide students on their journey?

The genetic *a priori*

- Students have a number of Piagetian ‘genetic’ or psychological *a priori* modes of thought, dispositions, expectations, implicit taxonomy or anticipations (Piaget, 1950).
- This approach to understanding our ‘default’ conceptual spaces is an evolutionary interpretation of Kant’s categories.
- Specifically, in physics, these conceptual spaces often correspond to what is known as ‘folk physics’.
- This approach is reliable in almost all everyday circumstances.

- The bad news: the genetic *a priori* does not save the evidence. It is often mistaken.
- For our purposes, focus on the difference between the average first-year student and a theoretical physicist.
- We desire that, after their journey, the student has the CS approximating those of a modern physicist.

- One answer is fairly simple:
  - we tell students where we started from (*folk physics*),
  - how we got here (*the entirety of the history of physics*),
  - and where we are now (*current physics*).
- This approach is the guided reenactment of the history of physics.
- Teaching is the imaginative reconstruction of the reasoning and experimental processes that lead to concept revision.

- Obvious downside, if given plenty of time: this path is as *uneconomical* as possible.
    - If we were to develop a fairly accurate model of the history of physics, it would be a dense directional graph that would take decades to understand.
    - It involves massive backtracking and unnecessary revision.
- Another downside, if time is limited: *incomprehensible*.
    - We cannot hold these minor distinctions in our heads in the amount of time available to the student.

Finding the ‘golden mean’

- Teachers want to minimise the number of CS between where the students begin their learning and the point at which the students can understand theoretical physics.
- The problem is ‘What are the fewest number of *manageable stages* of CS between models of “folk physics” and current theoretical physics that takes into account the introduction of prototypes (i.e. new evidence)?’

- We want to engage in concept-revision *when it is most economical*: we want the smallest number of necessary steps (where each preceding CS is most similar to the previous CS *and* simplest).
  - We can produce a history of science that is an idealisation of the research programmes in the history of physics.

Restating the problem

- The pedagogical problem is more appropriately stated as a balancing act between maximising long-run and short-term goals:
  - Long run: if these idealised research programmes are represented as nodes, what is the shortest path in a strongly connected directional graph *G*?
  - Short term: between any series of neighbouring nodes in *G*, which node preserves the structure of the vector space of the previous node while accounting for new prototypes?

- From Gärdenfors (2000), *we can model the similarities between nodes by distance*.
- An imperfect analogy:
  - We don't just want the shortest path to Rome; we want the shortest *and* 'safest' path to Rome, where safety is a measure of closeness between each city on the path.
  - Closer cities share customs, laws, language, currency, etc. than farther cities.

- We have very weak synchronic constraint on theories:
    - the predictions of theory do not contradict the currently accepted empirical evidence at some time *t* (Popper, 1959) (*NB*: the single realist constraint, referred to in what follows as *the rule of fit*) *and*
    - the theory is *closer to* the minimum message length when expressed in some language *L* than other available theories (Wallace and Boulton, 1968) (*NB*: this corresponds to Gärdenfors' naturalistic approach to simplicity, referred to in what follows as the *rule of simplicity*)

- We also have a very weak *diachronic constraint* on theories:
    - All else being equal, when given new empirical evidence that demands theory-change, prefer the theory that makes the fewest possible revisions to conceptual spaces (I will call this the *rule of preservation*).

- What is the shortest *distance* in a strongly connected directional graph?
- The answer to this question will maximise our short-term and long-term aims:
  - The shortest distance is the most economical *and* 'safest' path from our starting node to end node: it maximises both similarity between nodes and the simplicity of each node.

- This approach emphasises important diachronic constraints to movement between nodes. We have a far more richer and specific *theory of movement*.
- We have a way to model which key points should be covered in moving from pre-theoretical beliefs to a well-informed student:
    - The path corresponds to the key research programmes that *should* be taught to students, if our aim is to maximise both economy and 'safety' of their educational 'journey'.

These rule guide future movement

- I’ll cover three examples of the benefits of diachronic theory when applied to *future movement*:
- explaining why we accept the rule of simplicity,
- explaining why we accept the rule of preservation and
- solving one interesting version of the problem of underdetermination.

A problem justifying the rule of simplicity

- The rule of simplicity works well if we desire ease in acquiring new conceptual spaces—it guides our reconstruction of our *past* movement in the most economical and cognitively safe way.
- Why accept simplicity as a rule of motion in future, unguided inquiry? Why should we think that the simpler theory is more likely to be true?
  - The obvious answer: we shouldn't. The simplicity of a theory T does not give a reason to believe that theory T is true or likely to be true (Kelly, 2004, 2007).
  - Synchronic approaches aren't so helpful here.

A diachronic justification for simplicity

- From a diachronic view, simplicity is a good *rule of motion*: start simple, follow the available evidence, and engage in minor concept-revision.
    - If there is a deformation of CS that violates Voroni tessellation, shift to a nearby simpler theory that equally fits the available evidence.
    - Why? We've proceeded down a path of a series of nodes in concept-space that leads to complexity.
    - But what diachronic reasons to we have to prefer simplicity over complexity in our theories besides a violation of Voroni tessellation?

- Kevin Kelly (2004) argues on the basis of formal models, a simplicity preference is part of a procedure that reliably approaches the truth given the available evidence *via* the fewest dramatic changes of the conceptual space *en route*.
- Kelly speaks of developing methods that over time minimise the number of retractions by making each step in theory-preference non-arbitrary by picking the simplest theory (or conceptual space that corresponds to the theory), given the available evidence.

- What is *not* needed is an assumption that simplicity or preservation of theories must ‘point to' the truth in the short-term. This approach is question-begging.
- What *is* needed is the minimisation of U-turns, detours or ‘scenic routes’, which we can see in the history of science.
- If we can ‘prune’ these dead-end branches before they grow too large, we preemptively end unproductive inquiry.

- ‘…disregarding Ockham’s advice opens you to a needless, extra U-turn or reversal in opinion prior to all the reversals that even the best of methods would have to perform if the same answer were true. So you ought to heed Ockham’s advice. Simplicity doesn’t indicate the truth, but it minimizes reversals along the way’ (Kelly, 2004, 492)
- Kelly’s approach isn’t exactly novel: it is explicitly linked to Putnam’s (1965) *n-trial predicates* and computational learning theorists’ *mind-changes* (Jain et al. 1999), and makes frequent reference to Gärdenfors’ work (Kelly, 2004, 2007).

- So long as we want the most economical path:
  - When two theories presently equally fit the available evidence, preferring a more complex theory will do nothing in the long run but produce extra steps.
  - Preferring the simplest available theory will require the shortest number of steps.
  - The simplest available theoretical web is expressible as a Voronoi tessellation.

A problem justifying the rule of preservation

- *The rule of preservation*: do not alter or discard theories unless there is a reason to do so. Quine calls this principle the 'maxim of minimum mutilation' (1990 p.14).
- If our goal is to believe some theoretical web that is true, then it is hard to see why we should accept the rule of preservation.
- Why should we privilege *our* theoretical web over another? Because it is ours? Why not switch from one to another or remain indifferent?
  - Again, synchronic approaches aren't so helpful.

- But a diachronic approach has a helpful answer:
  - We don’t take seriously rival theories that we are BIVs, subject to Evil Demons or in the Matrix.
  - We don’t seriously consider whether Aristotle’s physics is correct anymore, even if it could be salvaged.
  - Why? They’re intellectual ‘dead ends’—a waste of time.
- We’re permitted to shift when a rival theoretical web has made a case worth listening to, e.g. Everett interpretations.

- This solution is ‘pragmatic’, rather than ‘epistemic’, only if we assume that the solution *must* be synchronic.
- Diachronic grounds help guide decision-making over a long period of time, minimising paths that will likely not be fruitful.

The problem of underdetermination

- A related problem to justifying the rule of preservation is a formulation of an *interesting* version of the problem of underdetermination (rather than versions that are closer to Evil Demon or BIV puzzles):
    - for any body of evidence there will always be more than one scientific theory that can accommodate it (a modification of Psillos 1999, 164).
- Why should we privilege one theory over the other? Given that the body of evidence accommodates both theories, shouldn't we remain indifferent?

- Rather than seeing these versions of underdetermination leading to pessimism, we see there is a positive side if we focus more on diachronic justification. The corollary of the problem of transient underdetermination:
  - for any comparison between theories, some possible body of empirical evidence will discriminate between the two theories.

- This new accepted available evidence helps us decide which theory to accept by following the *rule of fit*:
  - don’t accept theories that contradict currently accepted empirical evidence.
- If there isn’t currently accepted available evidence that doesn’t discriminate between two theories*, search for the evidence*.
  - (*NB*: the problem of articulating which contexts acceptance of empirical evidence by the scientific community is rationally held will not be addressed here).

A guide for future theory-preference

- If T2 fits the accepted available evidence better than T1, prefer T2 (*rule of fit*).
- If T2* is simpler than T1*, and there is no access to a body of evidence that discriminates between T2* and T1*, prefer T2* (*rule of simplicity*).

- The weakest rule of the three is the *rule of preservation*: if T2′ and T1′ fit the available empirical evidence and are equally simple, but T2′ is more similar to previously accepted theory T0′ than T1′, prefer T2′
- *unless* T1′ remains viable after criticism. Then choose whichever one.

- In summation, the computational and naturalistic approaches may provide *joint* grounds for our preference for simplicity in explaining our past inquiry in the sciences, how we guide inquiry in the classroom, and our future inquiry when faced with theory-choice—
- When faced with a problem with a theory (e.g. refuting evidence), *all else being equal,* we have epistemic grounds for our choice of the simplest theoretical web that fits the available empirical evidence and preserves the most of our background assumptions and auxiliary hypotheses.

Bibliography

Brown, R. 1991. The Laboratory of the Mind: Thought-Experiments in the Natural Sciences. Rutledge.

Gärdenfors, P. 2000. *Conceptual Spaces: The Geometry of Thought*, MIT Press, Cambridge, MA

Gärdenfors, P. & Frank Zenker, 2014. 'Communication, Rationality, and Conceptual Changes in Scientific Theories', in *Applications of Conceptual Spaces: The Case for Geometric Knowledge Representation*. Springer

Jain, S., Osherson, D., Royer, J. and Sharma, A. 1999. *Systems That Learn: An Introduction to Learning Theory*. Camabridga: M.I.T. Press.

Kelly, K. 2004. ‘Justification as Truth-Finding Efficiency: How Ockham's Razor Works.’ *Minds and Machines* 14: 485-505.

Kelly, K. 2007. ‘Ockham’s Razor, Truth, and Information’

Popper, K. 1959. *The Logic of Scientific Discovery*. Rutledge.

Piaget, J. 1950. *The Psychology of Intelligence*. Rutledge & Kegan Paul.
Putnam, H. 1965. “Trial and Error Predicates and a Solution to a Problem of Mostowski,” *Journal of Symbolic Logic* 30: 49-57.

Psillos, S. 1999. *Scientific Realism: How Science Tracks Truth*. London: Routledge.

Wallace, C. & Boulton. 1968. ‘An information measure for classification’, Computer Journal, Vol. 11, No. 2

Waddington, C.H. 1954. ‘Evolution and Epistemology’, *Nature*, Vol. 173.