David Lewis and the Kangaroo: Graphing philosophical progress

BENJ HELLIE

Penultimate version: 25 February 2016

Philosophy's Future: The Problem of Philosophical Progress, edited by Russell Blackford and Damien Broderick. New York: Blackwell, 2016.

Abstract: Data-driven historiography of philosophy looks to objective modeling tools for illumination of the propagation of influence. While the system of David Lewis (1941–2001), the most influential philosopher of our time, raises historiographic puzzles to stymie conventional analytic methods, it proves amenable to data-driven analysis. A striking result is that Lewis only becomes the metaphysician of current legend following the midpoint of his career: his initial project is to frame a descriptive science of mind and meaning; the transition to metaphysics is a rhetorically breathtaking escape from this program's (inevitable) collapse. Understanding this process both aids a more focused debate whether it counts as progress, and also presents novel affordances for partisans on both sides to learn from Lewis's right and wrong steps.

Keywords: David Lewis, data-driven history of philosophy, history of recent analytic philosophy, metaphysics, network analysis, philosophy of mind and language

Introduction

If there is progress in philosophy, it comes about through *progression* in philosophy: through development and change over time in what philosophers write. Without a clear understanding of philosophical progression, of how what philosophers write has actually developed and changed, it is premature to ask whether there is progress in philosophy. It would be sending the jury out before the trial starts.

The trouble is that we probably don't know how philosophy develops and changes. Professional philosophers are in the best position to say, because

philosophy is hard, and we are the experts. But because we have skin in the game, it is hard to be impartial: this is especially true if the issue is philosophy *now*, or in the recent past.

Perhaps there are unexplored ways of taking the personal element out. Like the scientists, we could work up some models we think are liable to convert the mess of data into comprehensible and reasonable answers to our questions, gather the data, and run the models on it. Even if the answer that pops out shakes us up, we won't be able to shoot the messenger.

This chapter describes the current state of an ongoing research project employing a data-driven, impartial, model-building approach to history of philosophy. The project starts small, focusing on the work of a single philosopher—but not *too* small, because my subject, David Lewis, an American philosopher active from 1966 to his untimely death in 2001, is the philosopher with the greatest (direct) influence on recent philosophy (Healy 2013b).

Lewis's publications number around 130 (depending on how you count: S. Lewis 2015)[1], and are almost invariably of high quality, with groundbreaking works in most areas of philosophy. Lewis was both extraordinarily creative and extremely careful, with a mathematical facility unparalleled in the field; his technical contributions form the basis for much of the work done in the formal analysis of meaning, and his contributions to logic and decision theory place him in a position of high influence in those fields. Much in the Lewis corpus remains still to be mined: work in philosophy of language that had lain fallow since the early 1980s is now beginning to re-energize the field, as we finally catch up with Lewis.

While he rarely repeated himself, and despite his protestation against having set out to be a systematic philosopher (Lewis 1983a), there is a great deal of interconnection throughout his corpus, making it a true "system." Despite his immense productivity, Lewis's corpus is largely consistent, and acknowledged changes of mind are rare; still, there is enough in the way of tension (and oddball views, and enigmatically compressed remarks, and unacknowledged changes of mind—perhaps even outright inconsistency) to seed the interest of the historian.

The diversity and complexity of Lewis's corpus makes it a challenge to get a handle on, considered as a whole. How is it structured, if at all? What is relevant to what? Does it have fundamental hypotheses, or is it an assemblage of largely disconnected parts? Are there discernible temporal phases, among which there is

[1] In a citation to David Lewis, the letter following the year represents the ordinal position, in S. Lewis 2015, among those from that year of the work cited (so 1979e = "Attitudes *de dicto* and *de se*").

discernible variation in overall approach? Can such variation be explained in terms of dialectical pressures? —Such questions are not easy to approach, when dealing with a philosopher who, in just one banner year (1979) published papers on *commands*, the *prisoner's dilemma*, the *direction of time*, the *common knowledge we use in conversation*, *what you learn when you look at your watch*—all of them groundbreaking classics—and, to top it off, on an alleged refutation of the contention that computers can simulate us (and Lewis didn't slow down much in 1980 or 1981, either).

Largely out of recognition of my inadequacy to the task of locating structural joints in Lewis's corpus, unaided, I have attempted to automate the task—to take myself out of the question of where the joints are, and return only to puzzle out their significance. My approach has been to construct *force-directed graphs* of Lewis's *autocitations*: citations of his own work (nodes are publications, edges joining them represent mutual relevance, the spatial arrangement works out on its own when nodes try to get away from each other but are constrained by edges). These represent his own in-the-moment judgments about what in his work is pertinent to what. Although Lewis made a number of pronouncements about what he had been up to (inter alia: Lewis 1983a, 1986b, 1994a), charting autocitations arguably provides a better guide to his *genuine* motivations.[2]

I initially set out on this task in an attempt to figure out what led Lewis to commit a certain (glaring) error in the 1979 "Attitudes *de dicto* and *de se*" (Lewis 1979e: more below) and then concealed the error from him (except for a brief moment of anxiety: Lewis 1996a n6; contrast Lewis 2001a) for the rest of his career. I think I found the answer, but I also learned a great deal more.

Lewis begins his career in the mid 1960s with a deep and rigorous understanding of a technical approach to representing information with "possible worlds," and a detailed and methodical strategy for using it to explain mind and meaning. The approach (to my mind) would turn out to be inadequate to the task (the big problem is a treatment of other minds by analogy to "black boxes"—which neglects the "first-person viewpoint," and thereby neglects the sharp distinction between our understanding of the mind and of everything else). This is foreshadowed in a small inconsistency between two streams in which Lewis

---

[2] The use of force-directed visualizations of citation-network graphs as a means to represent relationships of intellectual influence is lifted wholesale from Healy 2013a; Healy in turn is "basically copying" Neal Caren's prior application of this approach to the field of sociology (http://nealcaren.web.unc.edu/a-sociology-citation-network/). Each of us produces visualizations with Mike Bostock's D3 JavaScript library: https://d3js.org/.

pursues distinct questions about language (a resolution is proposed in Lewis 1980c); that aside, the overall development of the program is marked by superhuman rigor and thoroughness, and is structured into a legible progression of conceptual chunks.

By the mid 1970s, Lewis has wrapped the program up. In search of stuff to think about, he methodically scrutinizes each chunk in search of potential problems. Those he finds do penetrate, as I see it, to the core of his program, but (disappointingly) he reacts largely by evasion. A flurry of renewed activity starting in 1979 comes with the explicit recognition (at long last) that the first-person perspective is a problem for his view; here, too, Lewis combs through the conceptual chunks of his program looking for fixes and revisions. Unfortunately, this requires so much tampering that by the time it is done, the program has lost its coherence.

At this point, Lewis does something quite remarkable. He is known today as having been centrally concerned with *metaphysics*, the "pursuit" of the "elements of being," as he would put it, in his "self-blurb" for the photographer Steve Pyke (Pyke 1995). That does not square well with the first half of his career, when he was concerned instead with understanding mind and meaning in terms of the mathematics of information. Instead, 1983 marks a break of continuity, when he turns away from the earlier program and takes up a serious interest in metaphysics.

The kind of metaphysician Lewis defines himself as turns out to be significantly conditioned by the availability of resources during the earlier period that escaped the intense scrutiny he had given his hypotheses on mind and meaning. In light of a tangential connection of one of these resources to the earlier program, that previous program ends up recast, first, as a "proof of concept" for an encompassing metaphysical system of prior significance—its struggles everywhere do not amount to the collapse of a system for understanding meaning and mind, but to the inevitable growing pains of a system of metaphysics—and in light of the pervasive involvement in it of "possible worlds", second, as an application in an "indispensability argument" for that device in *On the Plurality of Worlds* (Lewis 1986a). In short, Lewis redefines himself in the 1980s as a metaphysician in order to transcend the collapse of his central program.

This transition has its pluses and minuses. On the minus side, when Lewis abandons the earlier program he muddies its waters, making it harder to see where it went wrong and therefore how to fix it. On the plus side, *Plurality* would become one of the two most influential philosophical works of our time (Healy 2013b). Whether this counts as progress depends on how one rates Lewis-style metaphysics

relative to the study of meaning and mind (as a philosopher of mind and language, I see it as the robots taking over the factory for their own ends); but Lewis's powerful metaphysical work has since inspired decades of vibrant discussion.

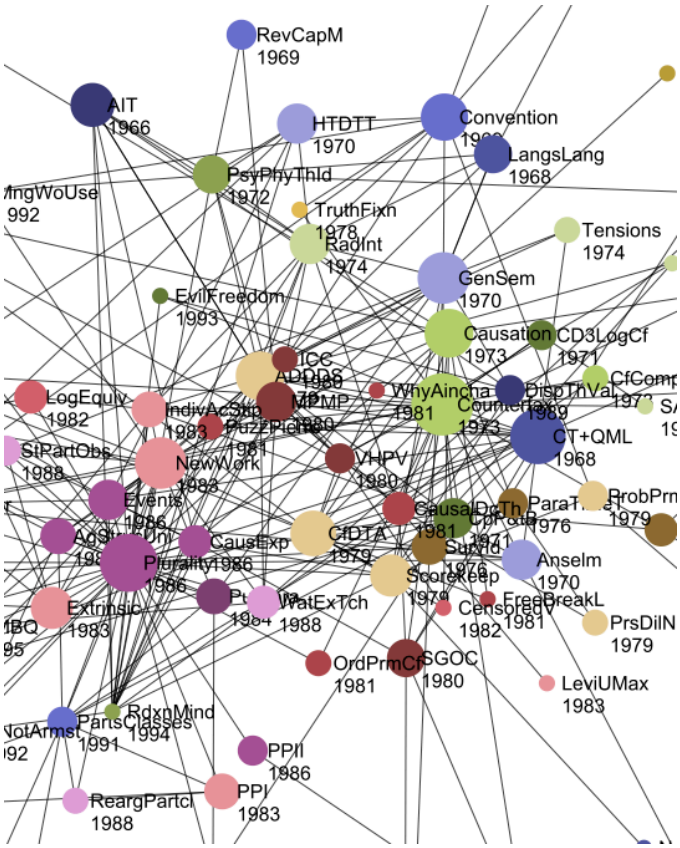
Graphing Lewis: coarse structure

My database of Lewis's autocitations (compiled by hand) covers 129 works published by 2014, each of them assigned a date.[3] Of these, 99 *systemic works* either cite or are cited by one another: there are 270 episodes of autocitation in all. With the assistance of David Balcarras, I have constructed force-directed visualizations of network graphs extracted from the data (Hellie 2016).

These are of two kinds: graphs of *development* and of *subject-matter*. Development graphs depict the raw data of what cites what, and when. So edges are "directed": there is a meaningful difference between the "source" of an edge and its "target"—namely, the source is the cit*ing* work, the target the cit*ed* work. In the visualizations published online (Hellie 2016), hovering your cursor over a node turns the outgoing edges blue, incoming edges red. I attempt also to incorporate temporal structure by making edges "want" to have a length proportional to difference in age between source and target. More influential works are more firmly anchored to their year, while the less influential are displaced.

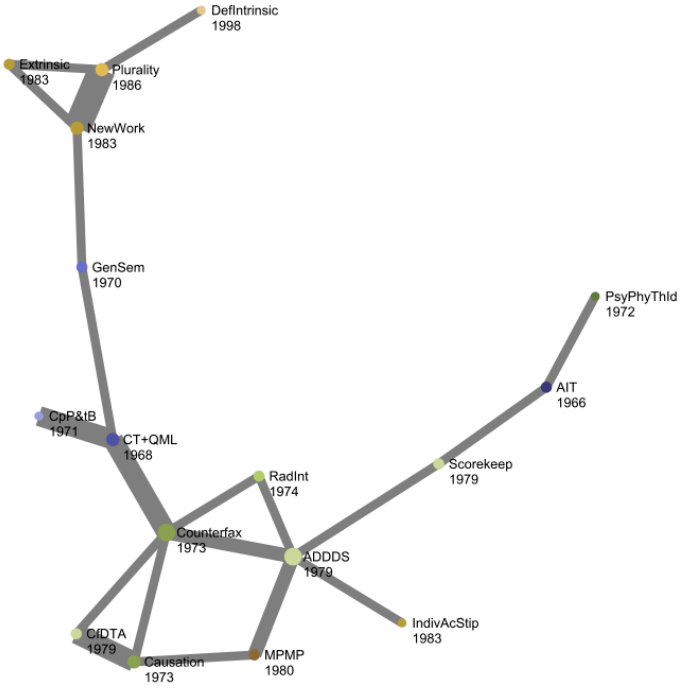
This suggests the use of development graphs to distinguish "cresting" periods when the work is of more systematic influence, from "waning" periods, with work of less systematic influence: cresting periods should have nodes that anchor the temporal order, while nodes from waning periods should get out of the way. The full-career development graph does in fact highlight three "bands" with many nodes clustering together in the temporal order (separated by "bands" without such clustering): 1966–74, 1979–81, and 1983–86. This makes it plausible that work from these three periods has a distinctive systematic importance in Lewis's philosophy. (For the record, this squares very easily with my subjective impression.)

[3] In all cases but two, this is the publication date. The exceptions are two works published significantly later than their composition, each of which I date to 1972: 1999a23, a 1972 handout; and 1975a, a subtly differing "noncanonical" earlier version of which had appeared starting in 1972 in various formats and locations.

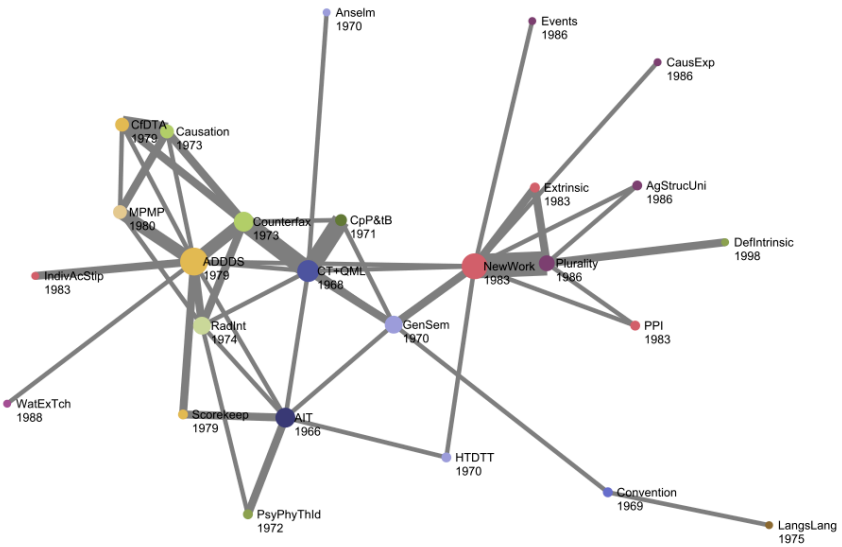


Subject-matter graphs are based on the more abstract relationship of the *co-citation*. When *A* and *B* are both cited by *X*, this is a "directionless" relationship; this is also the case when *k* or more works cite both *A* and *B*. Subject-matter graphs therefore offer an image of conceptual order purified of the temporal order. To reveal the conceptual backbone of the system, *k* should be set as high as is useful, in light of the general prevalence of co-citation in the corpus.

With $k = 5$ or above, the graph falls apart: the largest useful value is $k = 4$. The resulting graph of Lewis's work is the Kangaroo (displayed below), containing the 17 works inhabiting pairs co-cited four or more times. The model predicts that these 17 should approximate the works of most fundamental significance to Lewis's system; going by my intuitive sense, this is not far wrong. A further attractive feature of the Kangaroo is evident upon inspection: mostly spread out along a line, with a central "body" of interconnected works, a "head" at one end, and a pair of "leg" and "arm" spurs, this suggests an ordering of works across a subject-matter dimension, with more extensive interlinkage among a small nucleus of first-tier works (among the most central of which, note, is "Attitudes *de dicto* and *de se*": Lewis 1979e).



The distribution of the three "cresting" periods across the Kangaroo is strikingly uneven: the "head" (with its "ear") consists of four works following the 1983 onset of the third cresting period, with "New work for a theory of universals" (Lewis 1983e) alone in contact with older work; the remainder of the graph is entirely from the first two cresting periods, 1966–74 and 1979–81—specifically, with a "torso" and "lower tail" from the first, separated by a "hip region" of works from the second. This suggests that the third cresting period involved the taking up by Lewis of a significantly novel direction of research. That suggestion is strengthened at a slightly lower level of abstraction: the subject-matter graph for $k = 3$ depicts "New work" as a bottleneck between a network of only later works and a network of (with one exception) only earlier works.



I found this to be extremely striking. The sociologist Kieran Healy, in his study of philosophical influence from which my method is adapted (Healy 2013a), maintains that Lewis's "main area" is "metaphysics" (Healy 2013b). This correctly records Lewis's own later explicit self-estimation—in the late 1990s, Lewis explained his work to the photographic portraitist Steve Pyke by proclaiming himself "an old-fashioned analytic metaphysician" (Pyke 1995)—and (in light of the centrality to contemporary discussion Healy establishes for the metaphysical book, *On the Plurality of Worlds*: Lewis 1986a) the implicit view of the philosophical community.

But the disjuncture between the pre- and post-"New work"-Lewis suggests a more complicated narrative, on which Lewis *became* a metaphysician in the second half of his career, but spent the first half doing something else less easily categorized. After all, "New work," itself a metaphysical work, also divides those works, in the Kangaroo's head, that are clearly metaphysical from a broader array of works dealing with a broader array of topics: in the neck and arm, works on logic and linguistic meaning; in the upper torso, works on causation and "conditionals" (if–then statements); in the lower torso and tail, works almost all on the mind (with Lewis 1979d, dealing with linguistic meaning, interposed).

So perhaps Lewis *became* a metaphysician in 1983, with "New work," after starting out doing something else—something that involved a range of different areas, and which spread across both the 1966–74 and 1979–81 "cresting" periods.

Rise and fall of the black box mind

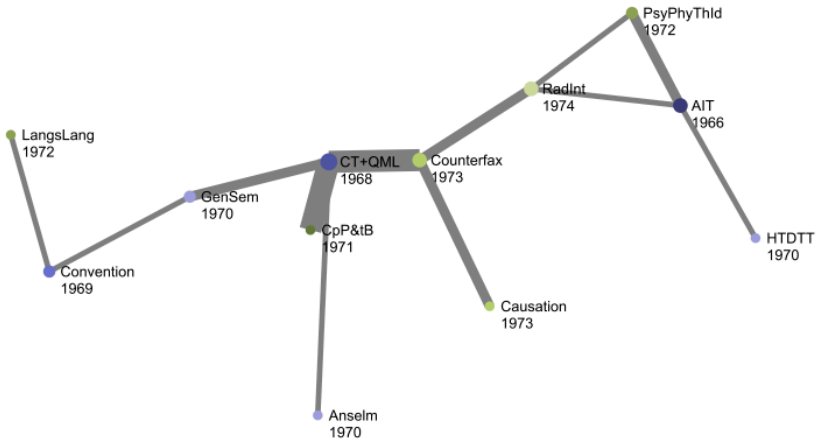
But if so, why? What was the earlier program concerned to do? And why, in the early 1980s, between the second and third cresting periods, did Lewis decide to do something else instead? —And could this have anything to do with what goes wrong in "Attitudes *de dicto* and *de se*"?

It occurred to me that the answer might have something to do with the longish waning period from 1974 to 1979, between the first and second cresting periods. It is not too hard to imagine the following sort of process: by 1974, Lewis has wrapped up the program with which he started his career. He then spends several years thinking about where to go next. That has been settled by 1979, but rather than then forging out on an entirely new program, Lewis instead revisits the old issues from a point of view he had not initially accommodated. Such renewed attention would be hard to motivate, unless in the waning period Lewis had taken note of some blind spot that undermined the initial development of the program. But if so, the second cresting period would be in the business of reworking the treatment of the initial bouquet of issues, to accommodate that blind spot.

If this narrative is correct, we should expect to see a conceptual ordering to the works up to 1974 that is mirrored in the works from 1974 to 1979 and then again in the works from 1979 to 1981. And we should expect also to see a dialectical progression, in which works to 1974 advance a series of proposals to fill out this conceptual ordering, works from 1974 to 1979 attack those proposals, and works from 1979 to 1981 respond to those attacks. Names for the periods suggestive of such a dialectical structure will be useful: I call them the *Programmatic*, *Critical*, and *Revisionary* periods, respectively.

The conceptual order of the Programmatic period can't be read straight off the Kangaroo, which falls apart without post-Programmatic works. The next option would be to look to the subject-matter graph for $k = 3$, with post-Programmatic works removed—but here we see the opposite problem, of too much interlinkage (works on logic link directly to works on mind, when the Kangaroo had put them at a remove). I split the difference by "hand-thinning" the second of these to remove all weight-3 edges that take shortcuts through the Kangaroo. The result depicts a structure of conceptual distances among twelve Programmatic-period works; sticking with the marsupial theme, I label it the Platypus.[4]

[4] Sadly, it is missing a leg.



At this point, network analysis has exhausted its capacities, and continuing the story requires some attention to the *content* of the works. After all, what the concepts *are* that are involved in those works is not a structural question; and while that question is relatively shallow and easily answered, the deeper question regarding the overall motivating spirit of the program is not an easy one at all.

Regarding the shallow question, the answer (roughly, but sharp enough for our purposes) is this: the works with which Lewis executes the core of his program fall into six categories. The two at the tip of the Platypus's tail (category I) deal with *how human language gets its meaning*; the next three (to the mid-torso: category II) with *what meaning is like, in general*; the two in the front torso (III) with *what conditionals mean* and *what causation is*. The head and bill are a bit less orderly in conceptual structure (because the basic work, Lewis 1966a, is about an *argument* rather than a *theory*, and the argument spreads across a few different issues). The three works in the head deal with several different issues about the mind. The work at the neck, 1974d (IV), deals with *what rationality is like, in general* (the other two, category V, are more concerned with *sensations*), while all three deal with *how we think about the mind*. Finally, the last three works (the bill—category VI—along with category V) deal with *abstraction*. To save space while providing illustrative detail, I will restrict consideration to *mentality*,

categories IV and V; work in progress discusses similar goings-on elsewhere in the system.

Regarding the deep question, the animating idea of Lewis's program (worked out in category II) is roughly that meaning is exhausted by "objective," viewpoint-free hypotheses to distinguish possibilities for the world. The idea percolates to categories IV and V as follows. (IV) *Rationality* involves doing various things with objective information (or misinformation): in particular, *opinion*—in the jargon, "belief"—is the possession of objective information. Belief about *the mind* (IV–V), therefore, is also understood in terms of objective information—in particular, using a striking analogy between humans and programmable bike locks to illustrate his view, Lewis proposes that we think of other minds as "black boxes": states of belief and sensation are unknown internal states, conceived of in terms of how they causally interact with sensory stimulation and bodily movement.

Turning now to the systematic Critical-period works, each of them can be assigned to one of the categories as raising a challenge for its core Programmatic hypothesis. In category (IV–V), Lewis begins to confront the issue of the *first-person perspective* (bike locks have no "inner life," no distinctive understanding of their own state, but quite clearly that kind of understanding is so important to our conception of mind that our understanding of bike locks is a completely inappropriate model). The Critical period only sidles up to it, in a pair of papers treating *time travel* (Lewis 1976b) and *brain-splitting* (Lewis 1976d). While these are cases where the first-person and objective views start to break apart, Lewis unfortunately overlooks this ramification, with these papers addressing instead technical issues in how to describe the objective facts.

Works from the Revisionary period also fit this categorization. In categories IV–V, Lewis at last squarely confronts the first-person perspective. Unfortunately, he does so through "internalism": the black box model remains true, but the box is understood just by the undulations it gives off at the surface; how it interacts with its environment is of no fundamental relevance. (IV) "Attitudes *de dicto* and *de se*" addresses the kind of viewpoint-sensitive information acquired by looking at a clock. (V) The delightful "Mad pain and Martian pain" (Lewis 1980b) pushes internalism one step further, by maintaining that talk about pain has in mind sometimes what it *does* (making people shout) and sometimes what it *is* (some sort of brain activity). These go together whenever the person is a normal human, but can be teased apart for nonhumans ("Martians," where we focus on what pain *does*) and for abnormal humans ("madmen," where we focus on what pain *is*).

By the end of the Revisionary period, Lewis's views are everywhere beset by instability. (IV–V) The black box model of our thought about psychology is implausible on its face, and internalism makes it even worse: a much superior theory has me understanding Fred through "empathy," by conforming myself to how Fred seems to be. (IV) "Attitudes", recall, contains a glaring error Lewis would never notice: the bizarre prediction that, if at 1PM I think it is 1PM and at 2PM I think it is 2PM, this should be treated as a "change of mind," of a piece with cases in which I recognize an earlier error—say, by going from thinking that goats eat cans to thinking that goats do not eat cans.[5] (V) As Lewis would eventually realize (Lewis 1995a), if the first-person perspective is really about what pain *is*, then if what pain *is* is a certain kind of brain activity, the first-person perspective should reveal that brain activity: unfortunately, "making discoveries in neurophysiology is not so easy!" Lewis handles this by rejecting our conception of mentality as a myth; unfortunately, he thinks pain is brain activity only because he thinks our conception of mentality demands it (e.g., Lewis 1966a, 1972b, 1994a): so I doubt Lewis's view can be given a stable description.

Lewis never acknowledges any of these concerns, so it will never be known whether they somehow added up to a nagging suspicion that something had gone wrong. For my part, I find it hard to imagine Lewis being oblivious to all of this tension. Nevertheless, the issue that officially motivates the transition to metaphysics is something else entirely.

A prolonged campaign in hindsight

The core idea of "New work"—Lewis 1983d, with which Lewis dons the mantle of metaphysics[6]—is anodyne. Start with an age-old question in the philosophy of science. Scientific theories begin with the actual but extend to cover nonactual possibilities; accordingly, any sensible theory can be exchanged for a theory compatible with the facts, but completely crazy about the possibilities. But then why think the sensible one better than the crazy one? More convenient, more plausible, sure; but what does convenience or plausibility have to do with truth? (Those of an "idealistic" cast of mind say "that's a bad question," but Lewis allies himself with the "realists.") Lewis borrows apparatus (the "universals") from his friend David Armstrong, renames it *naturalness*, and impresses it into answering

[5] This is the mistaken prediction that set this project in motion: see Arntzenius 2003 for details.

[6] "New work" is given pride of place as the leadoff paper in Lewis 1999a.

the question (the "new work"): the kind of possibility science cares about is constrained by what is most *natural*, while sensible theories characterize the *natural* and crazy theories the *unnatural*. This goes for any theory, including theories of other minds. In particular, if I interpret Fred's sentence "snow is white" to mean that snow is white rather than that snow is white just if things are exactly as they actually are and otherwise that grass is pink, my more natural interpretation is superior because it follows the relevant constraints on possibility.

But soon afterward (Lewis 1984a), Lewis makes a delicate yet immense adjustment to the relation between *naturalness* and meaning. Recall that he has now been for some time an internalist: I do not get to help myself to Fred's interactions with snow and whiteness in interpreting his sentence "snow is white," but only to the course of undulations undergone by Fred's body: a course of undulations that doesn't have anything in particular to do with snow or whiteness. But with this minimal basis, why bring snow or whiteness into the story at all? All sorts of crazy meanings are compatible with just those undulations. Internalism is evidently false, right? Lewis rejoins, in effect, that if *naturalness* can get rid of crazy interpretations in general, it can do the same here. Problem solved! (Well, no, to my mind: now *naturalness* is no longer in the business of protecting good theories "out in the wild" against crazy skeptics, but of serving as a deus ex machina to protect a tottering piece of philosophy against a compelling worry.)

Here we come to a happy accident. The Platypus paper in category VI is "How to define theoretical terms" (Lewis 1970b). That paper is about abstraction: washing out the specific subject-matter of a theory and leaving behind its less specific structure. The facts on the ground combine with this structure to return the specificity to the theory. (Its role in the Programmatic period was to explain how black-box mental concepts link up to the brain activity inside the box, as part of a treatment of the mind–body problem.) Lewis assumes that an abstract structure washing out all connections to the world outside is what is in common between the beliefs of everyone who undulates just as Fred does. It is the facts about how things are around Fred that combine with this structure to restore its specificity.

The conceptual link established between *naturalness* and this approach to abstraction generalizes: among all the ways to restore detail to abstraction, *naturalness* breaks ties. In its use to rescue internalism, this link perhaps does not make for particularly satisfying theory; but soon after, it will make for excellent rhetoric. Thanks to the conceptual link with the newly discovered *naturalness*, the

formerly peripheral device for abstraction can be now recast as having been central all along.[7]

A further pair of happy accidents provides further articulation to the image of the natural. First, the Critical period discussion of brain-splitting (Lewis 1976d), recall, overlooked the interesting question about the first person: it instead became involved in the logical properties of one object dividing into two (amoeba-style). Lewis's treatment appeals to "stages": material objects existing only for an instant. Second, the Programmatic discussion of causation (Lewis 1973a,b), primarily for logical reasons, offers a glimmer of its reduction to matters of fact, stripped of any consideration of power or powerlessness. Bundle these together, and the image appears of a "mosaic world," in which the most basic components are connected to one another only very "weakly," by distance in space and time—an image perhaps affirmed by David Hume.

Taken together, these components make for a program of revealing all apparently strong connection to be an artifact of our concepts. In the preface to his second volume of collected papers (Lewis 1986b), Lewis announces that his prior research "seem[s] to me in hindsight to fall into place within a prolonged campaign on behalf of the thesis I call "Humean supervenience." [A]ll there is to the world is […] just one little thing and then another.[…] All else supervenes on"—is an abstraction from—"that." Regarding earlier struggles, "There is room for endless argument over the details, but I remain confident that at every step mentioned the connection is something like what I have said—enough like it, anyway, to allow the cumulative Humean supervenience of one thing after another." The shakiness of the original program on every question of initial interest is no longer a problem. Details, details: the objective was never understanding mind and meaning, but rather "a priori reductionism about everything" (Lewis 1994a) to the Humean mosaic of natural properties.

Lewis's self-explanation to Steve Pyke, mentioned above, continues by highlighting the two strands in his metaphysical program:

> I am an old fashioned analytic metaphysician, in pursuit of hypotheses about what things are the elements of being, and about how all else may be reduced to patterns of these elements. I am notorious for claiming that these elements must include many that are merely possible, no part of this world that we ourselves live in, but none the worse for that. (Pyke 1995)

---

[7] In the $k = 3$ graph, "New work" interposes itself between the category II and VI papers.

The first strand here, of course, is Humean supervenience.

The second strand breathes life back into a further region of unblemished work from the Kangaroo: the logical works in category II, in particular "Anselm and actuality" (Lewis 1970a), in which Lewis proposes "modal realism," the thesis for which he proclaims his "notoriety," to the effect that all possible worlds are just as real as this one. In the 1981-composed preface to the first volume of papers (Lewis 1983a), this material is regarded as an encumbrance; but defense of modal realism soon fills the immensely influential 1986 book, *On the Plurality of Worlds*.

The earlier program makes an appearance in *Plurality*, where discussion of its central phenomena is compressed into one section: modal realism is argued to be "indispensable" in their philosophical analysis (as well as that of several other phenomena). The issues of basic significance for the earlier program are merely instrumental to the goals of the later program.

Whether it is *progress*, this upending of the relative significance of the original target of analysis (meaning and mind) and its enabling theoretical apparatus (abstraction, individuation, possible worlds), is an evaluative matter—significantly, a matter of taste. Those drawn toward "realism" will praise Lewis's generalization beyond his earlier interests to more abstract structures in reality, and the deepening of theoretical investigation into those structures thereby enabled. Those drawn instead toward "idealism" will see Lewis's initial interests as having all along been the proper target of philosophical attention; will think of the collapse of the early program as the inevitable comeuppance of a misplaced realism regarding these matters; and will regard the subsequent metaphysical program as a distraction from the real issue raised by this collapse—namely, how to discriminate the good and bad in the program, and then fix it.

But a point of agreement, I hope, will be that network analysis has promise in guiding first-order philosophical investigation toward progress. Realists may welcome the conceptual archeology of Lewis's adventitious arrival at his peculiar metaphysical outlook: perhaps as a pointer to remaining work in Lewis's program, or to alternative programs; perhaps as a solvent for arguably deadlocked debate (Wilson, this volume). And we idealists who regard David Lewis as a model theoretician may be happy to learn that the *structure* of Lewis's early program was sound—and that, but for the program's realist "*stuffing*," it would have worked.

References

Arntzenius, Frank 2003. Self-locating beliefs, reflection, conditionalization, and Dutch books. *Journal of Philosophy* 100:356–70.

Healy, Kieran 2013a. A cocitation network for philosophy, http://kieranhealy.org /blog/archives/2013/06/18/a-co-citation-network-for-philosophy/, 18 June 2013.

———— 2013b. Lewis and the women, http://kieranhealy.org/blog/archives /2013/06/19/lewis-and-the-women/, 19 June 2013.

Hellie, Benj 2016. Archive of visualizations for "David Lewis and the Kangaroo", http://bit.ly/dklkanga.

Lewis, Stephanie R. 2015. Bibliography of the work of David Lewis. In *A Companion to David Lewis*, edited by Barry Loewer and Jonathan Schaffer, 562–71. New York: Wiley.

Pyke, Steve 1995. *Philosophers*. Zelda Cheatle Press.