6

Modality

Contents

The notion of *modality* is almost inextricably intertwined with metaphysics, some kind of theory of what is real, what exists, and why (a theory of 'first causes'). At the center of the commonsensical theory is the real world, but the idea is that there exist, or at least there can exist, other worlds. This idea is most clearly supported by the commonsense notion that the world existed yesterday and will exist tomorrow, even if it will be slightly different from what it is like today. In 3.2 we already discussed that the current world V_n requires two modal versions: V_b for the past and V_a for the future, and in 6.1 we will considerably refine this idea by describing a a nonstandard theory of the standard temporal modalities.

A central question of metaphysics is the existence of the things that are co-present in all worlds, the things that do not change. Are these collected in another world? Are there other worlds to begin with, and if there are, in what sense are they real? In 6.2 we use the same technique to introduce an ideal world V_d where rules are kept, and investigate the real world in relation to this. In 6.3 we use an even simpler technique to bring epistemic modality in scope, and in 6.4 we reanalyze defaults.

6.1 Tense and aspect

In 3.2 we introduced the naive theory of time, and described how it requires at least one, and possibly two somewhat imperfect copies of V_n to explicate word meaning. When we say that statements about these subworlds, and especially statements that involve more than one of these, have modal import, we rely on the broad family of theories known collectively as modal logic. (For a terse discussion, see S19:3.7, for a book-length one see Blackburn, Rijke, and Venema, 2001.)

A. Kornai, *Vector Semantics*, Cognitive Technologies, https://doi.org/10.1007/978-981-19-5607-2_6

Fig 6.1 depicts part of a two-way infinite graph, the nodes of which correspond to (full or partially specified) *possible worlds* and the directed edges depict the *accessibility relation* that obtains among these. The left-to-right direction corresponds to the flow of time, with nodes to the left (right) of the vertical line in the past (resp. future), and the vertical line itself marking *now*. The bulleted node is the real world, as of now, those above and below are alternative worlds at the same time. There is only limited movement across the horizontal timelines: not every past world is compatible with the present world, and not every future world is accessible from it.

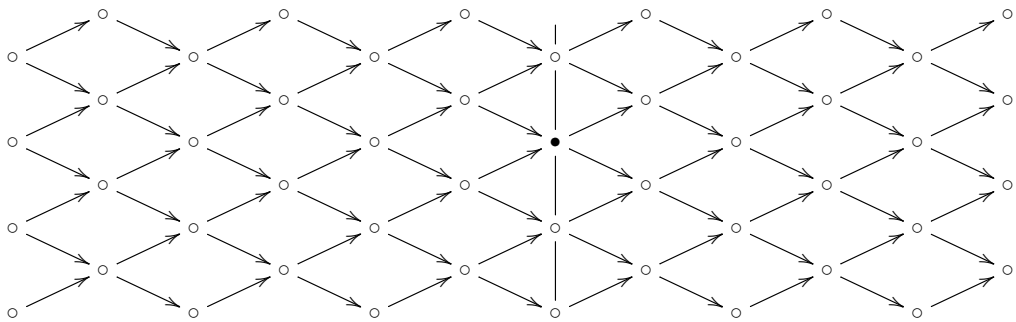

Fig. 6.1: Modal accessibility

The sophisticated reader can find all kinds of faults with Fig. 6.1. How can we have a single (absolute) time that flows synchronously in all worlds? How do we know there are alternative timelines, rather than a single, fully deterministic one? What makes us think we could move across timelines? In response, let us reiterate the disclaimer we already made at the beginning of Chapter 3: linguistics and cognitive science, our primary tools here, are highly unlikely to contribute to contemporary cosmology, precisely because in physics we already assume things such as a continuous timeline (possibly with a singularity at the beginning) which go beyond the scope of the naive theory. No matter how much the study of the naive system contributes to a better understanding of early natural philosophy, especially the Presocratics, Zeno in particular, there is no reason to suppose that it can contribute to contemporary natural philosophy (physics, chemistry, biology, and so on) since the naive system is already insufficient to sustain integer arithmetic, let alone real numbers, functions, and higher mathematical constructs essential to the practice of science.

In fact there is no reason to believe that the commonsense theory is in any way trivial. A fuller mathematical reconstruction of the temporal aspects requires more sophisticated tools from mathematical logic and analysis than are generally employed in mathematical physics. (Readers who wish to refresh the standard modal concepts may find S19 Ex. 3.17-19 helpful.) Time instances are best seen as infinitesimal neighborhoods (called

monads in nonstandard analysis, and *timelets* in smooth analysis by Bell, 1988). Such a formulation goes a long way towards capturing some of the seemingly paradoxical properties of time instances: that they are discrete (monads corresponding to different time instances have no overlap), that they have no tangible beginning or end, that they have nonzero length yet ordinary time is suspended with a single timelet. For reasons that we will discuss shortly, we consider Lawvere-style smooth analysis, as opposed to Robinson-style nonstandard analysis, to offer a better formulation for the dynamic aspects of commonsense temporality.

Since normally it is verbs that carry tense marking, the expectation is that it will chiefly be verbs that require explicit temporal clauses. Perhaps surprisingly, temporal modality is already required to treat suffixes like `-th` (*half, third*, and *quarter* are irregular and not discussed here) which we must define by some unit that is getting divided (or broken up, as in Skt. *bhinna*): `-th -ad -ias -ta part, in whole, before(divide)`. To conceptualize a fraction we must assume a preexisting unit. Again, this does not imply, or even suggest, that quarks, with fractional charges, must have originated by breaking electrons up – what the evidence shows is that the *words* for fractions post-date the words for integers, not that fermions post-date bosons. `-th`

A more typical entry, with extension in all three temporal slices, is *pause*, defined as `lack action, before(action), after(action)`. Similar to fractions, which are tied to their units, pauses are inescapably tied to some action that is being paused. There are many relational nouns that only make sense only if some other entity is invoked, but these are typically in the same V_n temporal slice, whereas the act of pausing requires the presence of activity both before and after. A particularly interesting case is provided by *through*, which we analyze as an adposition of fictive motion: `before(=agt on side), in =pat, =pat has side, after(=agt on other(side)), =pat has side[other]`. To understand this word, we need to invoke a full story about its object, a `body` (in the sense of 3.1) with two sides, and virtual movement that starts and ends outside this body, but is inside for a period. `pause` `through`

To build a more refined formal language describing such cases, we introduce further notation, with π_i defined as the projection of the polytope in V_i, for $i = b, n, a$. Since the default is `now` (event time), this is left unmarked in the lexicon, where only `before` and `after` are overtly marked. Typically these three projections are identical, especially for nouns, where we expect the realizations π_i to be isomorphic in the spaces V_i. Only 15.5% of the defining vocabulary contains overt reference to `before/after/cause_`.

It is worth emphasizing, particularly to the reader familiar with the standard theory discussed in S19:3.7, that projecting from the joint polytope (or system of polytopes) in $V_b \times V_n \times V_a$ to one of the components V_b, V_n, or V_a is *not* the same as the extension of a concept in the past, present, or future. Take a stable noun such as *food* : `material, gen eat`. It is quite possible (and historically common) for various things that were not considered food to become generally accepted as food, and conversely, for materials that were earlier considered suitable for eating to get off the menu. But such changes `food`

are slow, adiabatic, and do not affect V_b or V_a which are best thought of as *sharing the timelet* with V_n.

Within the adiabatic approximation `food` is an eternal noun, a proper member of the deontic world V_d that we will discuss in 6.2. This is true even though it is a commonsensical law of nature that food is perishable, i.e. no *instance* of food lasts long (though special efforts to salt/smoke or otherwise preserve it may extend its usable lifetime). This is one point where the modern theory, capable of distinguishing between *instances* and generics, is arguably superior to the naive theory. That said, we can still add a rule *food is perishable* by nonstandard methods without triggering a contradiction.

Since *perishable* means 'likely to decay quickly' (LDOCE) and `decay` is given as `change[slow], after(lack health)`, we need to consider the timescales in more detail. When we say the magnetic field of the Earth decays, this is on a hundred thousand year or longer timescale, and can only be measured with sophisticated instruments. But even for ordinary decay processes from decaying buildings to decaying teeth a multi-year timescale is implied. When the LDOCE posits quick decay the implication is that for food, the process is quick relative to the multiyear timescale inherent in decay, just as the *enormous flea* discussed in 4.3 that is enormous only on a flea scale. As a practical matter, the timescale is weeks, possibly days or even a few hours, but certainly not minutes or seconds as we have with ordinary action verbs.

On the near-instantaneous timescale embodied in `before/after`, food does not change, or changes only with imperceptible speed. This is something well captured by smooth analysis, which reconstructs derivatives with the Kock-Lawvere Axiom:

> If D is the timelet around 0, and f is any function from D to $\mathbb{R}$, there is a unique real number a such that for all $d \in D$ we have $f(d) = f(0) + d \cdot a$

The axiom guarantees that within a single timelet *all* functions are linear 'linelets' with a unique tangent a. This lets us define derivatives at every instant t as the unique real number $f'(t)$ that satisfies $f(t+d) = f(t) + d \cdot f'(t)$ for any d in the same timelet as t. Notice that there are no higher derivatives that could be obtained by closer inspection of an infinitesimal neighborhood, in fact it is a characteristic property of smooth analysis, as opposed to nonstandard analysis, that infinitesimals d (called *minims* in this theory) satisfy $d^2 = 0$. What we have in smooth analysis is a theory restricted to continuous functions, embodying the famous Leibnizian principle of *natura non saltum facit* 'there is no jump in nature'. To handle the case of food perishing, all we need to add is that the process is akin to the movement of the hour hand on a clock, possessing a derivative, but one that is too small to be perceptible. It is not that the derivative itself is a minim, it is measurable by instruments of sufficient precision to be an ordinary non-infinitesimal quantity, it's just that our senses do not offer this level of accuracy.

Where does this leave us in regards to pauses? The commonsensical, if ahistorical, answer is that we equate movement with nonzero derivative and conversely, we equate zero derivative with lack of motion. This is perhaps best illustrated by a ball bouncing against a hard surface: the idealization that distance from this surface is measured by $|t|$ is

simply not available, since at $t = 0$ this function only possesses a left derivative -1 and a right derivative $+1$. Rather, we have to assume that either the ball or the surface is not entirely rigid, that for the time of the impact the center of the ball actually approaches, and departs from, the surface in a smooth fashion. This function will have zero derivative at $t = 0$ from either side. Therefore, the ball is `pausing` at the surface.

The geometrical picture associated with smooth analysis includes not just a clear picture of derivatives as the tangent function of the linelet angle, but also the idea that the entire curve is built from such linelets, just as a circle can be conceptualized as an infinite-degree polygon. The reader interested in how the entire apparatus of multivariate calculus can be built on smooth analysis is referred to (Bell, 2008). Here our goals are more modest: we concentrate on linking the naive theory embodied in language to early theories of natural philosophy, and consider the task of reconstructing modern physics entirely out of scope. A key element of the medieval theory of dynamics is the notion of impetus, what today we would define as speed times mass. It is impetus that endows objects with an intention to keep moving in the direction that they are already moving in. The use of impetus resolves the Aristotelian quandary of why a rock, once thrown, does not fall to the ground as soon as the hand is no longer supporting it, but rather follows a parabolic trajectory.

This much, while clearly insufficient for planetary motion (which will have to wait until Newton) is quite sufficient for Oresme, Buridan, and the great scholastic thinkers whose line actually goes back to Aristotle (via very significant Arab contributions that go well beyond mere transmission and commentary). To the extent that the picture provided by smooth analysis is highly intuitive, we begin to see the intellectual leap that separates second derivatives from linelets. Our intuition, grounded both in everyday experience and in linguistic cues, may readily supply the idea of local linearity (sometimes called *micro-straightness*), but this does not extend to second derivatives. Indeed, if two points of a curve are in the same timelet, they are also on the same linelet, so to build second derivatives we need to reify the derivative as a function on its own. (The very idea of time-distance diagrams goes only back to Oresme, Aristotle didn't have these at his disposal.)

The first person to wrestle with the issue was Zeno, and his paradoxes demonstrate quite clearly that certain commonsensical assumptions about time and space, if held jointly, will result in contradictions. The linguistic conception of time is discrete, but this immediately leads to the paradox of Dichotomy: a discrete system of instances cannot be dense, with a halfway point between any two instances. The standard solution is that time instances are indeed point-like, but we have infinitely many – nonstandard analysis endows these points with monads surrounding them while being disjoint from one another. In Robinson's version, monads have rich internal structure, in Lawvere's version, they are just tiny lines that can be characterized by their centerpoint (position) and direction (impetus). That we need some kind of continuous time for conceptual semantics is evident: in S19:3.3 we wrote

> The key temporal notion in the prolepsis is not so much the idea of *time* itself as the idea of a *process*. It seems that humans (and in all likelihood, all mammals) are endowed with a perceptual mechanism that inevitably makes them perceive certain sensory inputs as processes. Try as we might, we cannot perceive the flight of the arrow as a series of states: what we see is a continuous process. The compulsion to do so is so strong that even truly discrete sequences of inputs, such as frames of a movie, will be perceived as continuous, as long as the frame rate is reasonably high, say 20/sec.

More debatable is the concept of space, whether we see it as composed of small discrete voxels, or as a continuum. The egocentric coordinate system we discussed in 3.1 is actually neutral on the issue, presenting space as being composed from a few discrete regions like *inside* and *outside* but without any implication that movement within a single region is imperceptible. But Zeno's argument against the reification of space is still worthy of consideration: if everything has a place, what is the place of place? If there are voxels, can they be occupied by (parts of) physical bodies? If we suppose they can, where are the voxels themselves going to be? Again, it is the background assumptions behind the paradox that matter for us: clearly, Zeno is making the commonsensical assumptions that (i) *two things cannot occupy the same space at the same time* and (ii) *a thing cannot be at two places at the same time*. We will discuss each in turn.

`place` We begin our analysis by defining *place* as `point, gen at`. This is an atomic, point-like entity 'where things can be' as opposed to the conceptual `{place}` schema we defined in 3.1, but it makes little difference whether we consider just the point or the voxel, the 3-dimensional monad surrounding it: the import of (i) is still that no two objects can occupy it. This is considerably stronger than what our primitive negation element, `lack`, is capable of expressing, since `lack` is most natural in situations where some default expectation is not met (see 4.2), whereas in (i) we wish to express an absolute negative. How is this negative enforced? Everyday experience shows that this is done by one object either forcing the other one out, or by not letting the other one in.

Since an exception is provided by 'shapeless' objects such as liquids and gases which, according to everyday experience, can in fact mix in the same place, instead of objects we will insist on *solids*, both in the `4lang` sense of `firm/2215` and in the original LDOCE sense of 'having no holes or spaces inside'. The requisite sense of solidity is easily defined by the uroboros vocabulary as `lack empty(place/2326) in`. (This will also help with firm objects like Matryoshka dolls which can in fact occupy the same place as far as their center of gravity is concerned.) This notion of solidity, actually quite close to the idealization of the *convex rigid body* that we rely on in classical mechanics, is sufficient for restating principle (i) as

```
{solid at place, other(solid) at place} cause_ move   (6.1)
```

Eq. 6.1 is a far cry from modern dynamics, where impulse is conserved in each dimension, so the resulting movement can be computed with great accuracy: here what

moves, the object, the other object, or perhaps both, is left underspecified. The reference to `move` in Eq. 6.1 makes clear that it is a constraint over different time instances that is responsible for (i), making what appears to be a static principle into a principle of (proto)dynamics. Principle (ii), to which we turn now, will be different: no movement is implied.

Since everyday (solid) objects extend over a volume, they actually *can* be at two places at once: the bridge is at the left side of the river, and also at the right side. To go further, we therefore need to restrict principle (ii) to point-like objects and to point-like places. This departs from the general sense of *place* which permits overlapping or even containment: the tree is in the forest, the forest is in the country, so a single point, say the point of an arrowhead, can be both in the tree and in the country. Since `4lang` defines *place* as `point, gen at` and *point* as `place, lack part_of` we get the restriction both to point-like objects and point-like places for free. The knowledge engineer would probably state principle (ii) as

`place`
`point`

$$\texttt{point}_1\ \texttt{at}\ \texttt{point}_2,\ \texttt{point}_1\ \texttt{at}\ \texttt{point}_3 \Rightarrow \texttt{point}_2 = \texttt{point}_3 \qquad (6.2)$$

While very familiar to contemporary thinking, Eq. 6.2 is much farther than we are prepared to go. Several notational conventions used in Eq. 6.2, though typical of the uniqueness statements used elsewhere, are beyond our formal language. First, the main connective $\Rightarrow$. We never offered a non-causal theory of implication, and in fact deduction in a system that carries smooth analysis is of necessity a weaker, intuitionistic type (Moerdijk and Reyes, 1991). Second, and perhaps more important, we don't have a theory of indexing, as will be clear for readers of 3.3. Finally, the equality symbol '=', while close to negating our primitive `other`, is subtly different (more powerful). `4lang` makes a distinction between `equal/191 azonos idem identyczny` and `equal/565 egyenlo3 aequalis ro1wny` and being `other` really means 'non idem' rather than 'non aequus'. To appreciate the distinction, consider Pappus' proof of Euclid I Prop 5 known as *pons asinorum*: given a triangle ABC with equal sides AB and AC, the angles at B and C must also be equal. Pappus simply considers the triangles ABC and ACB: in didactical terms, he 'lifts up' ABC, 'flips it over' to make ACB, and lays it down on the original. This kind of subtle leveraging the notions of identity and equality bothers many people precisely because of (ii): how can the same triangle be in two places at once?

Staying within the limitations of `4lang` , we need to say `gen lack {thing at place(two)}`. The number *two* is defined as `number, one in, other in, two follow one`. Here we are free to ignore the ordinal aspect, which is irrelevant for our reconstruction of (ii):

`two`

```
gen lack {point at point, point at other point}
```
(6.3)

The syntactic mechanism will automatically identify the first three occurrences of `point` in Eq. 6.3, but not the third and the fourth since this is precisely what `other` means (see 1.6). One way to think about this is to consider principle (ii) part of the meaning of being *at* some location.

It is worth emphasizing that a trivial extension of (ii), "n things cannot be at $n+1$ places at the same time", is impossible to formulate within the limitations of `4lang`. It is of course quite possible to build better resourced systems, but sooner or later these will also run out of steam. An interesting case is provided by the obverse of (ii), the Pigeonhole Principle, which asserts that $n+1$ objects cannot be placed in n boxes without putting at least two in the same box. Proving this using only a polynomial size proof (the total number or symbols of the formulae in the sequence) and keeping each formula at constant depth (unlimited fan-in) is simply not possible (Ajtai, 1994).

Here the question is not just why we cannot have n objects in $n+1$ places at the same time, but rather why this general truth is hard to demonstrate already for $n=1$? Eq. 6.3 highlights the difficulty, not being `at place(other)`. Since *other* is defined as `different`, which in turn was defined in 4.2 with the aid of Leibniz' Principle of Indiscernibles, it is true for any `X` that it cannot unify with `other X`. This is guaranteed not so much by the semantics of unification (where overriding certain values, especially defaults, is possible) as by the semantics of *other* – this is what being other means. The `other` schema depicted in Fig. 1.3 demands the coercion (as discussed in 3.3) of `X` and `other X` to `one` and `other` respectively. To see this word in action, consider *passenger* `person, person[travel], person in vehicle, other(person) drive vehicle`.

other

passenger

There is, perhaps, a more general lesson here: it is the lexical meaning of core elements, such as `other`, that drives the large-scale behavior of complex operations, such as unification, and not the other way around. Given that specific lexical behavior cannot be deduced from general principles, we may as well adopt a 'generators and defining relations' style of description.

6.2 The deontic world

Almost all (perhaps all) languages and cultures have some notion of another world, generally populated by powerful anthropomorphic beings, ranging from gods and angels to evil spirits and devils. The major exception is the

> kind of disembodied use of higher forces that is taken for granted in Chinese metaphysics: "Thus to say that 'High heaven shook with anger' by no means implies that there is a man up above who shakes with anger; it is simply that the principle (li) is like this [that is, that crime deserves anger]." (Graham, 1958), p24, cited in S19:3:10.

It is this kind of Heaven, populated by principles (eternal laws) rather than by ghosts and spirits, that is closest to the deist metaphysics common in occidental philosophy starting with Edward Herbert's De Veritate. Broadly speaking, we use the same methodology relying on common notions *(notitiae communes)* (innate, universal semantics) as all rationalists from Herbert onwards (at least according to Chomsky's 1966 presentation of rationalist thought), but we don't undertake to faithfully reconstruct any of these

systems here. In S19:3 we offered an automata-theoretic reconstruction of the kind of patterns (rules, regularities) that we may consider central both for elementary guidance of behavior 'don't immerse your hand in boiling water' and for stating laws of nature 'unsupported objects will fall'. Here our main interest is explicating such statements in vector semantics, and we start with Yi-ch'uan's example

crime deserves anger. (6.4)

For this purpose, it will be quite sufficient to define *crime* as `action, illegal` `crime`
and trace `illegal` through `bad for_ law` to obtain `action, bad for_ law`. `illegal`
In fact, by tracing further *bad* as `cause_ hurt` we end up with an even more compact `bad`
definition of crime: `action, hurt law` – this has the advantage that we don't have to get sidetracked with the issues of experiencer subjects (see 2.4) that the use of `for_` would bring in tow.

The modal element is from *deserve*, defined as `before(=agt DO <good>),` `deserve`
`should[after(=agt GET/1223 =pat)]` where *should* means not just an option, but the `right/1191` option, an idea that is central to the deontic modality. We could also analyze *anger* further as `feeling, bad, strong, aggressive`, but `anger`
this would not take us further toward our goal, which is to express the pattern at hand as a normative rule.

In the world of norms, it's not just that the right thing *should* happen, but that it *does* happen, `after(gen angry)`. It's not that there is a man above who shakes with anger, but rather that everybody, the generic subject, will. Putting all this together we obtain the `4lang` translation

`before(action hurt law), after(gen angry)` (6.5)

Let us pause for a moment and observe that we have done a great deal more than translating Eq. 6.4 to Eq. 6.5 in that the target semantics is quite universal: for example Hungarian *törvénysértés* 'crime' is a compound that would be literally translated as *law-hurting*, and since *hurt* is defined as `cause_ {=pat has pain}, offend` we `hurt`
also obtain a semantic reading for English *offence*, namely that it is the law that is getting offended.

Given the pre- and post-conditions in Eq. 6.5, we can recognize this as an instance of the `cause_` primitive discussed in 2.4, so we obtain

`(action hurt law) cause_ (gen angry)` (6.6)

This lends itself to further generalization on both sides. First, it is not just actions, but any form of hurt that causes anger, and second, `gen angry` is just a restatement of `anger`. Therefore, we obtain

`hurt cause_ anger` (6.7)

as a truly general principle that has Eq. 6.4 as a consequence. By the definition of `cause_` this amounts to the co-presence of `hurt` in V_b and `anger` in V_a or, in the two-state temporal model we have `hurt` in V_n and `anger` in V_o. (Observe that in such a model not only does hurt cause anger, but anger also causes hurt, reminiscent of the

unbreakable cycles of violence we see all too often in tribal societies, where new generations simply take up the grievances of their forbears and keep on fighting like the Hatfields and McCoys.)

It is precisely the lack of a long-term sustainable commonsense temporal model that leads to the metaphysical search for first causes. Commonsensical `cause_` is simply the existence of a cause in one temporal model (V_b or V_n as the case might be) and an effect in a subsequent model. Once this is coupled with some Parmenidean principle of *out of nothing nothing comes* we must continue the backward search for a sufficient reason: what caused the cause? What caused the cause of the cause? If the underlying temporal model is cyclic, as with Ecclesiastes, we end up with simple, irreducible cycles of eternal change: day causes night, night causes day, there is no beginning and no end.

The more general three-state model of past, present, and future depicted in Fig. 6.1 does not cycle back from future to past. Rather, the underlying commonsensical temporal model includes a notion of a (discrete) timeline with successive states (time instances), with *present* always being mapped on a single state, *past* as a half-line extending with the previous state, and *future* another half-line gradually consumed by the passing of time, much as a wheel rolling over a surface only touches a single point of the ground, leaving a half-line trace.

There is nothing in language that is probative about the precise nature of this timeline, and commonsensical reasoning must be replaced by some form of more structured philosophy to be able to reason about infinities. If the timeline is finite, at the first step we must suppose some self-necessitated being, or unmoved mover, giving us the Kalam cosmological argument, whose attraction lies precisely in the fact that it hews so closely to commonsensical notions.

`god`

This is not to say that any commonsensical system will *prove* the existence of god, but it is quite obvious that the existence of *god* or gods `being, religion has, has power, er_ nature` is compatible with common sense. As we have argued elsewhere (Kornai, 2010a), "once the names of major religious figures and the titles of sacred texts are treated as pointers to the encyclopedia, there remains nothing in the whole semantic field of Religion that is not definable in terms of non-religious primitives". In other words, religion is *possible*, but by no means *necessary* for a commonsensical worldview.

With this, we have arrived at a central problem of modal logic: what is possible, and what is necessary. Let us begin with a standard example of a non-existent being, the *unicorn* 'an imaginary animal like a white horse with a long straight horn growing on its head'. We know of no law of biology that would make unicorns impossible – after all, both white horses and animals with long straight horns exist. We may search for unicorns in the past (perhaps remains can be found) or in the future (perhaps genetic engineering will produce some), though not every world is accessible to such a search. Even under the most cautious definition of 'possible', we find things such as dinosaurs which we know to be possible, and at the same time know to be nonexistent in the real world. Mathematics furnishes many examples that we know to be absolutely impossible: no search in past

worlds or alternative timelines will ever produce an algorithm for squaring the circle by ruler and compass.

Altogether, non-existence has an absolute form, impossibility, and existence has a weaker form, possibility. It is therefore reasonable to look for an absolute form of existence, which we can call 'necessity'. Things that enjoy only a weaker form of existence, or what is the same, a weaker form of non-existence, are called *contingent*. Altogether, we can build a three-point scale of existence, with necessity at the top, impossibility at the bottom, and the contingents in the middle. As long as we assume the accessibility relation to be reflexive, everything that is real (here and now) is either possible or necessary, and everything that is not real (here and now) is either possible or impossible. Looking through the lexicon we see that having a name like *unicorn* is no guarantee of possibility, unless imaginary worlds are also accessible, in which case of course it is (see S19:5.6).

Having a proper name is no more a firm guarantee of possibility than being named by a common noun. Here we take the position that the entire linguistic subspace L is composed of possible things, be they part of the basis (defining words) or be they expressed by more complex expressions. Where does that put us in relation to linguistic expressions that are known to designate impossible things, such as squaring the circle? The answer is that such expressions are also rigid designators (they mean the same thing in every possible world) it's just that they cannot be realized. The fact that something can be expressed in language is no guarantee that it's true or real.

In the naive theory, the type distinction between *true* and *real* is absent (cf. *a true friend/a real friend*). To bring this in line with contemporary logic where only well-formed formulas can be true/false and only objects can be real (there is never any doubt about the reality of a formula) requires a special predicate EXISTS whose type-theoretic signature is from matters[1] to truth values. In `4lang` we define *real* and `exist` as synonyms: we have `real igazi verus prawdziwy 1126 A exist` and also `exist van exsto byc1 2587 V real`. Similarly, we define *true* as `igaz verus prawdziwy 1125 A fact` and *fact* as `te1ny factum fakt 2323 N has proof[exist]`, something for which proof exists.

```
real
exist
true
fact
```

What constitutes proof? Again we eschew the modern proof-theoretic statement in favor of the naive theory, where *proof* is given by `prove`, and is a conjunction `after(other(people) know =pat[true]), real ins_`. Whatever a proof is, it is something that convinces others that its object is true and, moreover, it is an instrument (most effective means) to truth/reality. The most immediate proof of existence is supplied by our senses, compared to which scientific proof is effective only to the extent we consider scientific instruments more reliable than our senses. While our contemporary worldview considers scientific evidence to be superior to our senses, it is worth noting that in the age of science, many people insist that the Earth is flat, preferring the direct sensory evidence to more sophisticated theories.

```
proof
prove
```

[1] Recall that we use 'matter' as a cover term for objects and events/relations

Bodily feeling, proprioception, is the ultimate reality. This extends not just to strong signals of joy and pain, but to much weaker afferent signals pertaining e.g. to the status of our extremities: if my leg fell asleep it means it went numb. The physiological mechanism (restricted circulation, lack of oxygen, blocking of nerve path, etc.) is relevant only to the extent that a reasonable person *must* conclude that this is what is happening, *no* other conclusion is available. It is precisely this lack of choice in terms of explanatory mechanism that makes torture such an effective weapon: pain is real, as is joy. If `4lang`

`pain` we have *pain* as `bad, sensation, injury cause_` and *injury* as `damage, body has`, so pain/injury is tightly coupled to bodily sensation.

All of this is not to deny the existence of higher sentiments of joy and pain such as moral triumph and outrage. The claim here is that such higher sentiments are directly modeled on the direct, bodily sentiments, even if they are stronger, as they can be, when people withstand torture in the service of higher moral ideals. Again we follow the commonsensical notions that this is a matter of willpower, and `will/132` is simply a

`want` synonym of *want* which is defined simply as `=agt[feel[=agt need]]`. We have already seen that *feel* is bodily feeling, proprioception, and `4lang` skirts the issue of differentiating between needs and wants. (There is a culturally widespread theory that the two are subtly different, what you want is not necessarily what you need, and that you can't always get the former.)

The Swadesh list is a rich source of concepts with high universal presence: it is hard to imagine a language that doesn't have a word for *river* or *mother*, and perhaps even harder to imagine a world without rivers or mothers. There are no doubt such worlds exist, e.g. Mars is one, but the commonsensical theory that we are trying to model here loses its grip over these. There are two views of the deontic world, populated by eternal entities and governed by eternal laws, what we will call the *large* and the *small* view. Under the large view, every world of Fig. 6.1 is part of the deontic world, it is merely the case that some of them are inaccessible from the real world. The central deontic world V_D is one where rules are kept, children have parents, sugar is white, caramel is brown, and so on. We will compare this to the *default world* in 6.4.

Under the small view, the only accessible ones are where the concepts are instantiated, making them truly rigid designators in a way proper nouns are not (see 8.1), in spite of the original Kripkean intent. Consider the louse. We can well imagine a world where lice are absent, in fact most of us desire to live in such a world. There is nothing in the definition 'a small insect that lives on the hair or skin of people or animals' that suggests that a world without lice is beyond our ken. A world without insects is a bit harder to imagine, certainly this would result in an unimaginable ecological catastrophe with the very survival of humanity in doubt. Compare this to *mother* 'parent, female' and notice that the absence of parents implies the absence of children (indeed, `4lang` defines

`parent` *parent* by `make child`) and a whole different idea of human life pattern than what we have now, with concomitant loss of sense-making concerning a very broad swath of human languages (as spoken until this science-fictional future becomes reality) and concerning human cultures.

The closer we get to the core (defining) vocabulary the more we see the necessity of the concepts they name for meaningful discussion of any sort. In the vectorial perspective we can express this observation by saying that the special predicate EXISTS must be true for the basis vectors, but not necessarily for the entire space they span. The small view assumes that only things expressible in terms of this basis can exist. Most of them, like unicorns and lice, exist only contingently, but some of them, like *mother*, are necessary, if not for maintaining human life, at least for making sense of human discourse. The question must be asked: how do concepts like *river,* a Swadesh near-universal, but a defined word in `4lang` , behave? We have `river folyo1 fluvius rzeka 848 N stream, has water, in valley` but the defining terms don't all appear in the core vocabulary: *stream* is `<water> flow` and *valley* is `land, low, between <mountain>, between <hill>`. Altogether, this gives 'water flows between hills or mountains' where *hill* can be eliminated in favor of `on land, high, mountain er_` to yield 'water flows on land between more high land'. Concepts of this kind are possible, but not necessary. Their necessity, such as it is, comes from the fact that their shared knowledge is a precondition of communication.

`river`

`valley`

`hill`

The large view also permits existents that are not definable in terms of the basis. We are very aware of such existents, e.g. Fourier series, but we have trouble conceptualizing them as being present in the same world where we find mothers. The problem is not that Fourier series are contingent: to the contrary, mathematicians are as convinced of their existence as they can be. The problem is that they don't obey ordinary laws of nature, e.g. they have no weight, color, energy, shape, position, etc. We may say, together with all mathematicians who find a realist ontology convenient, that such object exist in a Platonic world, one that is even accessible to the human mind, but the commonsensical theory, which is our object of inquiry in this volume, has no grip over such worlds.

Even so, the commonsensical view remains useful in understanding the commonsensical concepts of probability discussed in Chapter 5 and, remarkably, the concept of instrumentality. We begin with *prior probability* of a matter X, defined simply as the proportion of worlds within our experience (think of the 'reverse light cone' terminating in the bulleted world of Fig. 6.1) where X obtained. To some extent, this is shrouded in lack of information, but if $l(X) = 0$ this means that EXIST(X) is false in all prior worlds, and $l(X) = 6$ means it is true in all of them. *Current probability* is a more mysterious notion precisely because we are attempting to estimate the proportion of existence across the vertical line of Fig. 6.1, those worlds that share the same time but, at the same time, worlds we don't really have access to. Much better to deal with *future probability* or *rational expectation*, which measures the proportion of existent Xs among the accessible future worlds.

As we shall see in 6.4, if no effort is made to change the outcome, matters continue on their default path: if $l(X) = 5$, X will continue to obtain, and if $l(X) = 2$, X will continue to not obtain. Further, no amount of effort can change $l(X) = 0$ or $l(X) = 6$ outcomes. This latter case, referred to as an act of God in the Anglo-American legal tradition, is delimited precisely because no human act could prevent it (when $l = 6$)

or bring it about (when $l = 0$). The real scope of human intentions is in the middle, particularly at $l = 3$, the broad domain where things are neither particularly likely nor particularly unlikely.

While the real world may be deterministic, the commonsensical world is certainly not: chance plays a big role, things can go many ways, and a key part of the human condition is that we don't know, some would say we cannot know, key events from the future. Accurate prediction of rare events is the hallmark of science, and Thales' predicting the eclipse of 585 BCE (this is now disputed, but see Couprie, 2004) is often described as the beginnings of science, as opposed to common sense. Be it as it may, *John hoped to win the competition* is different from *John won the competition*, and clearly the difference is lodged in *hope*, which is defined as `desire, want, =agt think =pat[possible]`. Next, *desire* is defined as `feeling, want`, and *want* as `=agt[feel[=agt need]]`. After all these substitutions, *hope* is still a feeling (by the agent) of needing something, obviously still a modal, and by looking up *need* we are not getting any closer, since this is defined as `=agt want`. However, the condition on hope `=agt think =pat[possible]` is helpful, as *possible* is given by `gen allow, can/1246`. We defer the analysis of *can* to 6.4, because it is simply given as `<do>`, and it is the optionality of doing, marked by the `< >`, that does the real work there. However, we can continue with *allow*, which is defined as `=agt[lack[=agt stop =pat]]`. Here the agent is `gen`, so *possible* means nothing (can) stop `=pat`, the object of the matrix agent's hope.

hope
desire
want
need
possible
can
allow

It is at this point of the analysis that the nondeterministic world-view comes into play: the mere fact that something *can* happen (there is no general force stopping it) doesn't in any way imply that it *will* happen. Fortunately, we are not restricted to hoping: we can improve the odds. Compare *John hopes to win* and *After a year of relentless training, John hopes to win*, or compare *John hopes to keep the wolves at bay* to *With his rifle, John hopes to keep the wolves at bay*. Altogether, *instruments* are tools that improve the chances of the desired outcome. Hope is good, but being prepared is better. Using our naive probability model, we see the immediate successor worlds of the current real world (see the arrows starting from the bullet in Fig. 6.1) as containing a distinguished default state, but for $l(X) \neq 0,6$ also many other states where the desired outcome X obtains, and many where it does not. What we want to say is that the likeliness of the desired outcome is increased by instruments. Needless to say, it has to be an instrument fit for the purpose: *With his hacksaw, John hopes to keep the wolves at bay* is as dubious as *With his rifle, John hopes to trim the beam*.

We have *instrument* as `object, work ins_, gen use, has purpose, at hand`, and note that in the `ins_` relation the instrument need not be a physical object, cf. *John won by cheating*. The prototypical instrument is a hand-held device with a specific purpose and this, seemingly accidental, aspect will be relevant for 6.3: what is near us is epistemologically certain. Our sense of touch is considered more reliable than our vision, which in turn is more reliable than our hearing. Our most reliable sense is is our proprioception: if we *feel* something, this overrides our auditory and visual

instrument

perceptions. By definition, *feel* is `=pat in mind, =pat at body, =agt has body, =agt has mind`, something that brings body and mind together in an act of perception. To feel something means that it is something right here, within the body schema, that is being brought to mind. Unlike things we hear, and even things we see, what we feel is something that cannot be denied. feel

To summarize, instruments are simply goal-oriented likeliness-increasing devices. This again illustrates a point we already made at the end of 6.1: it is the lexical semantics of the elements such as the instrumental case marker `ins_`, defined as `=pat make =agt[easy]` that drives the way instruments are referred to in language, not some top-down theory (such as hierarchical ordering of thematic roles). This is not to say that conceptual definitions such as Fillmore's "The case of the inanimate force or object causally involved in the action or state identified by the verb" or Pāṇini's "most effective means" are useless. To the contrary, these are both powerful paraphrases for trying to get to the meaning of the instrumental marker, and for the analytically minded, they provide excellent guidance in trying to sort out what (if anything) can be considered an instrument in a given situation. Our own definitional attempt differs from these chiefly in being provided in a fully formalized language, in keeping with the overall plan of the work. ins_

6.3 Knowledge, belief, emotions

We now try to articulate some fundamental assumptions about knowledge and belief. First, these are things in the head. Gordon and Hobbs, 2017 trace back the standard Theory of Mind (ToM) to Heider and Simmel, 1944, and here we follow in their footsteps to the extent feasible, but concentrate on how ToM is reflected in `4lang` . In this particular case, the definition of as *thought* as `idea, in mind`, relies on two notions we will analyze further, *idea* and *mind*, but readers of 3.1 will know that the spatial `in` relation is used in earnest: the mind is a `{place}`, and thoughts are in it. This gets further specified by the longest definition in the entire core vocabulary: thought

```
mind tudat conscientia umysl1 2457 N
     human has, in brain, human has brain, think ins_,
     perceive ins_, emotion ins_, will ins_,
     memory ins_, imagination ins_
```

We will not do justice to the complex discussion that followed Premack and Woodruff, 1978 whether the the proper definition should include `animal has` rather than `human has`, but note that the tendency to typecast animals, machinery, and even simple household objects as 'having their own mind' is strong not just in children but adults as well.

We obtain our starting point, that thoughts are in the head, by transitivity of *in*: if thoughts are in the mind, the mind is in the brain, and the brain is in the head, thoughts are in the head. We use `idea` as a near-synonym of *thought*, defining it as `in mind, think make`. More interesting is the relation of the nominal *thought* and the transitive verb *think*, defined as idea

```
think gondol cogito mys1lec1 907 U
      =pat in mind, =agt has mind
```

There is a subtle intransitive/transitive alternation often seen in psych verbs: if *John thinks* it is not necessarily the case that he is thinking of something – anybody who has ever struggled with putting thoughts into words will recognize the cases when the object cannot be formed easily, or at all. In English, the object of thought is typically expressed in a prepositional phrase, the agent thinks *of* something, or *about* something. This can easily be encoded by `"of" mark_ =pat` or `"about" mark_ =pat`, but the cross-linguistic variability is such that we refrained from doing so.

Second, the thoughts in the head are ontologically just as well established as the objects/events/qualities in the real world. We follow Meinong (see Parsons, 1974 for a clear modern exposition) rather than Frege, who places thoughts in a 'second realm', the internal world of consciousness. We would like to strongly discourage the reader from thinking about this in New Age terms, how consciousness creates reality, etc. Rather, this is a straightforward explanation of the human capability to *model* all kinds of things, from alternative outcomes of actions (as required for weighing the fitness of instruments for this or that purpose) to predicting the behavior of other agents. Further, the evolutionary advantage conferred by modeling ability is overwhelming: in any competition for resources if A can model B but not conversely, A is far more likely to obtain the resource.

Third, the assumption of thoughts in the head being real inevitably leads to the conclusion that other things in the head, such as feelings, emotions, desires, ... must also be treated as real. This, of course, is everyday human experience, and the commonsensical theory of emotions views them as humors flowing through the body. To better articulate the commonsense theory we have already gone one step further, endowing feelings with direct, non-negotiable reality in proportion to the reliability of the sense that conveys them. At the top of this hierarchy stands proprioception, followed by touch, vision, smell, and hearing in this order. Thinking is generally considered less reliable than our senses, and this includes discounting our own thoughts in relation to the words of the sages. Whether we like it or not, this is precisely the advantage that traditionalism and revealed teachings have over rationalism.

At this point, the reader may wish to revisit the discussion of grammatical moods and logical modalities in S19:7.3, but for greater convenience we summarize the 4L logic approach used there, which relied on the introduction of two more truth values in addition to the standard T (true, $\top$) and F (false, $\bot$), called U (unknown) and D (unDecided). Negation, as standard, makes F out of T and T out of F. In 4L the negation of U is U, cf. Codd's 'missing data'. The modal operator K will mean *known*, or rather *learned*, and will be given by `after(T or F)`.

The other nonstandard truth value, D, maps out the scope of agentive decisions and free will in terms of `before` and `after`. At any given time, the truth of a statement may depend on our own decision. Tomorrow morning I may drink tea, or I may not; this matter X is unsettled in all theories of free will (except in the denialist version,

which takes all such matters to be deterministically set in advance). In 4L the negation of D is D: if I am undecided about something I must perforce be also undecided about its negation. D means a nondeterministic transition `after`, to T or F, but not to both, and in this regard it is not at all like the 'both' value of Belnap (1977). The operations $\neg$, $\wedge$, $\vee$ are defined by the truth tables given in Table 6.1 below.

	T	U	D	F
$\neg$	F	U	D	T

$\wedge$	T	U	D	F
T	T	U	D	F
U	U	U	D	F
D	D	D	D	F
F	F	F	F	F

$\vee$	T	U	D	F
T	T	T	T	T
U	T	U	D	U
D	T	D	D	D
F	T	U	D	F

Table 6.1: Boolean operations in 4L

In 6.4 we will refine this simple theory of decision-making with a key observation: matters cannot stay undecided forever, not making a decision generally amounts to making a definite choice of letting the default operate. Certainly, if I defer the decision whether to drink tea until noon, this is for any observer quite indistinguishable from having made the positive determination not to drink any in the morning. We will use the modal operator S to describe the process of *settling* on a decision, meaning `after(T or F)`, where *or* carries the full force of the logical primitive `"_ or _" mark_ choose.` `or`

How is, then, the modal operator K, the act of learning, different from the modal operator S, the act of decisionmaking, especially as both satisfy `after(T or F)`? The most salient difference is in the frequency of the outcomes: if no learning takes place, we generally assume positive statements to be false, whereas if no decision is made, we generally assume that the default will carry the day (be true). Since the everyday experience that we are surrounded by an ocean of falsity, with truth being a rare find, seems to extend even to scientific studies (Ioannidis, 2005), we see no need to argue the point about K in detail here, S will be discussed in 6.4.

Clearly, emotions are as real as other things in the head, and in fact electrocardiogram-based emotion recognition systems can reach remarkable accuracy (Hasnul et al., 2021). Subjectively there doesn't seem to be a significant difference between bodily sensations like *feeling hot* and emotions like *feeling angry*, and most of the `4lang` definitions for emotions eventually go back to *feel* `=pat in mind, =pat at body, =agt has body, =agt has mind`. This is true both for basic emotions listed in `4lang` such as *anger* `feeling, bad, strong, aggressive` or *desire* `feeling, want`, and for abstract categories such as *feeling* `mental, other cause_, joy is_a, sorrow is_a, fear is_a, anger is_a` and *emotion* `state/77, in mind, feeling.` `feel` `anger` `desire` `feeling` `emotion`

Furthermore, the same holds for the entire emotion vocabulary, very much including words not explicitly listed in `4lang` such as *grief* 'extreme sadness, especially because someone you love has died' (LDOCE); 'very great sadness, especially at the death of

someone' (Cambridge). To reduce this to the core system, we first note that *-ness*, a clear deverbal and deadjectival noun-forming suffix (ignoring lexicalized cases like *business*)
sad is not essential. We can use *sad*, already defined as `emotion, bad` in the definition of *grief* as `sad, <{=agt love person die} cause>`. For the 'extreme, very
sorrow great' part `4lang` actually offers *sorrow,* `emotion, ER sad`, suggesting a better definition such as `sorrow, <{=agt love person die} cause>`. The naive theory of emotions embedded in `4lang` is not very sophisticated, but the links between sorrow 'dolor' and badness `cause_ hurt` are laid bare.

As in other semantic fields (Buck, 1949 devotes an entire chapter to emotions), we resist the temptation to offer a full taxonomy. Many words that Buck considered key
pity are removed during the uroboros search, for example *pity* has the following definition: `sorrow, {other(person) suffer} cause_`, but we see no reason to trace these exhaustively, let alone to trace all emotionally loaded words one may wish to consider. Broadly speaking, the naive theory treats feelings along the Hippocrates/Galenus lines as vapors or liquids (humors) flowing through the body, and we see traces of this in the free use of several motion verbs with emotions as subjects *joy flooded him*, or *his blood boiled* etc. We offer a mechanism for uncovering such taxonomies by tracing the definitions to the core, but we do not offer a policy.

6.4 Defaults

Perhaps the cleanest statement of defaults comes from programming languages such as C++, where function arguments can be equipped with default values. Other familiar examples include standard unix/linux utilities like `ls`, which will, when invoked with a directory path argument, list the contents of the directory in question, but will list the contents of the current working directory if invoked with no argument.

Natural language offers many similar examples, where a default object is assumed if no overt object is specified. Often the default object is highly unspecified as in *eat* <*food*>, other times it is highly specific, as an *expect* <*give birth*>. On the unspecified end, we often find cognate objects as in *sing a song, think a thought,* ... and very weak subcategorization as in *prove* <*something*>. Neither of the extremes poses a great challenge to a modal treatment of defaults invoking the large deontic world V_D or the small deontic world V_d respectively. As a practical matter, over 6.3% of LDOCE definitions contain defaults encoded by the keyword *especially*, as in *admit* 'to say that you have done something wrong, especially something criminal' or *rat-a-tat* 'the sound of knocking, especially on a door'. `4lang` relies on defaults even more heavily, with 13.8% of the core definitions containing clauses demarcated by $\langle\,\rangle$ (Rule 6 in 1.6).

We begin with a simple case, where the ambiguity is caused by predicating the default. Consider *-ist* as in *pianist, receptionist, scientist, tourist, violinist*. There is something of a slippery slope between characterizing a person for whom the notion expressed by the stem is important, as in *Calvinist, Marxist, Unionist, abolitionist, activist,* ...,
-ist and naming a profession, as in *archivist, anaesthetist, artist,* .... `4lang` provides

person<profession>, think stem_[important], "_-ist" mark_ stem_.

In many cases, we are not sure whether the person is professional: *arsonist, balloonist, philanthropist, ...*. That the default is *profession* job, before(educate for_) is clear from the fact that in these cases we tend to treat them as such, e.g. we assume that the arsonist is a career criminal, the balloonist has undergone rigorous training and flies balloons for a living, etc. We even have a word, *amateur* whose main use is to defease this implication. In the cases where it is hard to distinguish professionals from amateurs, the default profession takes precedence over the more general person. profession

Profession descriptors are a subset of person descriptors (as long as we don't insist on strict InstanceOf typing, see 4.5), so the lexical rule for *-ist*-suffixation operates the same way as the more static entries we quoted from LDOCE above. More challenging are those cases where there are two, seemingly disjunct defaults, as in *bake* <*cake, bread*>; *drink* <*water, alcohol*>; or *can/1427* cylinder, metal, contain [<food>,<drink>]. Such entries resist the kind of analysis based on is_a, since neither *cake/bread* nor *food/drink* has a superordinate member that the subordinate (more specific) member could override. can/1427

To compare this to the Pāṇinian idea of "habitual, professional, or skilled" actors noted in 2.2, we need to analyze what the three-way disjunction between *habitual, professional,* and *skilled* amounts to. For *habitual*, LDOCE offers 'usual or typical' in one sense, and 'as a habit that you cannot stop' in another. The distinction is carried back to *habit* 'something that you do regularly or usually, often without thinking about it because you have done it so many times before' versus 'a strong physical need to keep taking a drug regularly'. It appears we can do away with the compulsive sense, especially as the formations where it is most prevalent *(chain-smoker, pill-popper)* are synthetic to begin with. This leaves something like 'usual, typical, regular, done without thinking, done may times before' for *habitual*. For *professional* LDOCE offers 'doing a job for money rather than just for fun' and 'a job that needs special education and training, such as a doctor, lawyer, or architect'. Finally, for *skilled* it provides 'has the training and experience'.

It seems quite hard to disentangle the senses of *professional* and *skilled* as both require the before(educate for_) aspect that we used in the definition of *profession*. In modern society, 'need special education and training' really means education/training that provides a license: practicing law, medicine, or architecture without a license is criminalized, no matter how skilled the practitioner. This means we can collapse the second and third terms of the Pāṇinian disjunction (no doubt distinguishable back in his day) to just *habitual* or *professional*, perhaps adding to the latter an optional default clause <licensed> which must be omitted for *employer, farmer, manager, ruler, waiter,* etc. at least until regulations are further tightened.

It is clear that the range of the remaining two options overlaps greatly, but perhaps differently for nominals obtained by deverbal zero-suffixation (a device we have no need for, given the type-free nature of 4lang); by deverbal *-er* suffixation; and by denomi-

nal *-ist* suffixation. More important, the identification of these sub-meanings is *post hoc,* relying on the subdirection (see 2.2) rather than on the parts themselves. A *habitual offender* is simply a person who has offended many times before, there is no implication that they get paid for it, or that offending required any education or training, let alone licensing. If *for our next outing, Jim will be the cook*, this does not make him a professional cook, or even a skilled one, just one who assumes the role, quite possibly without the benefit of special education or training. It is precisely because of the post hoc nature of the choice between the habitual and the professional reading that the rule lacks productivity: we don't have a notion of the *??habitual sleeper* not because nobody is trained in sleeping (some people with disorders actually are, but we don't consider them professionals for that) but because everybody is on the habitual branch of the definition of *sleeper, eater, breather,*... to begin with.

Returning to dual defaults, it is intuitively quite clear that we would want to follow Pāṇini and permit disjunction e.g. in *bake* `cook/825, =pat[<bread>, <cake>], =agt cause_ =pat[hard]` whose default object is either bread or cake, but not both. One way to resolve the issue would be the introduction of some abstract supercategory such as 'dough-based baked food' or 'victuals'. We call this the KR-style solution, as it is seen quite often in systems of Knowledge Representation. This is unattractive for most languages (cf. 5.3 for 'doors and windows' in Hungarian), especially as the first paraphrase sneaks in *bake* on the right-hand side of the definition, and the latter (together with its less current synonym 'aliments') defaults to food, whereas 'refreshments', at least in current usage, defaults to drinks. The KR-style solution also goes against the lexicographic principle of reductivity (see 1.2) that the definiens should be simpler than the definiendum.

`bake`

The key to the treatment of defaults is to see them as triggers for spreading activation. We will discuss the activation process in greater detail in 7.4, but the general picture should already be clear: if a default is present in a lexical entry, it is active unless it gets defeased. At the discourse level, such activity is easily tested by the immediate, felicitous availability of definite descriptions. Compare *I went to a wedding. The minister spoke harshly* (Kálmán, 1990) to *I went to a restaurant. #The minister spoke harshly.* The wedding script comes fully equipped with a slot for *minister*, but the restaurant script does not. In fact, one need not resort to the full conceptual apparatus of Schankian scripts or Fillmorean frames to see this, the lexical entry for *wedding* 'a marriage ceremony, especially one with a religious service' (LDOCE) already carries the religious service and its officator by default, whereas the entry for *restaurant* does not.

Under the view presented here, a *restaurant* is not fully defined by 'a place where you can buy and eat a meal' (LDOCE) because the same test *I went to a restaurant. The waiter spoke harshly* shows *waiter* to be available by definite description. The existence of a specific negative, *self-service restaurant* also points at the conclusion that waiters are present in restaurants by default, as are chefs, maître d's, busboys, tables, etc. The Oxford definition, 'a place where people pay to sit and eat meals that are cooked and served on the premises' shows the slots for cooks/chefs and servers/waiters, and *sit and eat* does

seem to imply the presence of a chair and a table. Whether the maître d' hôtel is a default feature of a restaurant seems very much culture-dependent, but a *real restaurant*, as we shall see in 7.2, can hardly do without.

Let us return to conjoined defaults. Consider *ash* `powder[<grey>, <white>, <black>], {<wood> burn} make`. What is the default color of ash? The word *ashen* suggests 'pale gray' but an *ashen face* 'looking very pale because you are ill, shocked, or frightened' (LDOCE) is actually not grey, just pale. The larger encyclopedia, https://en.wikipedia.org/wiki/Shades_of_gray is already overwhelming, and a broader search leads to sites such as https://simplicable.com/new/ash-color which call into question even the Knowledge Representation-style solution relying on some technical term (in this case, *grayscale*). ash

There are cases like *broadcast* `signal, <radio,television> receive` where a KR-style solution is easy. Unlike *aliments* discussed above, where the defining word is lexicographically unreasonable, here we could use *antenna*, not just as something common to TVs and radios, but also as the instrument of both broadcasting and reception. But there remain cases like *opponent* `person, oppose, <compete>, <in battle>` where the defaults are rather contradictory between friendly competition and adversarial battle. In the spreading activation model we don't have to make early choices between polysemous senses or pretend that these involve a single abstract sense. Rather, the system can resolve later on which of the adjacent polytopes is meant. broadcast opponent

We began S19:1.1 with two Fregean principles, the better known Compositionality, and that of *Contextuality*:

> Never ask for the meaning of a word in isolation, but only in the context of a sentence.

In computational linguistics, relating word senses to contexts is known as the problem of Word-Sense Disambiguation, see Agirre and Edmonds (2007) for the state of the art before 2010. Perhaps the greatest step forward in solving the WSD problem was the introduction of *dynamic* embeddings that produce a word vector based on context. Unfortunately, this is a black box solution, and part of our goal here is to understand the mechanism of disambiguation. Defaults, contradictory defaults in particular, offer an important insight into the structure of lexical entries: while the basic structure is conjunctive, their joint activation, by spreading, is disjunctive. The broad agentive `-er stem_-er is_a =agt, "_ -er" mark_ stem_` is simply 'one who stem-s' (cf. *buyer, sleeper,* ... rather than 'one who habitually stem-s', so there is no disjunction to consider. The more narrow agentive *-er*, and *-ist* are, perhaps just like in Sanskrit, ambiguous between the habitual and the professional readings (cf. *smoker, exhibitionist* for the former and *plumber, pianist* for the latter) but we see no supercategory that connects these two: rather, we see these as disjunctive by virtue of being defaults. The work is done by the `person[<profession>]` clause which defaults to `profession`. We have to do extra work to escape this conclusion in order to fall back on the default `person`, and this extra work is unrelated to any notion of habituality, since the pros -er

obviously 'do stem' habitually. In synthetic compounds *teetotaler, navelgazer,* ... we assume the work is done during the formation of the compound, in other cases we may have to bring in the compulsive aspect we chose to ignore above.

The entire network of lexical entries is remarkably tight. We have seen that from the uroboros core every word can be reached in three steps via the LDV and LDOCE. Three is the maximum: those familiar with the use of `4lang` can often write a one-step definition that relies only on the uroboros core. (By now, most readers will have seen enough examples and will understand the principles well enough to try themselves.) Since the average number of clauses within the V2 uroboros set is 2.66, if we let spreading proceed through any undirected 'associative' path, we may activate the entire vocabulary in 5-6 steps starting from the words of any sentence. Consider *colorless green ideas sleep furiously*. *Color* immediately activates `sensation, light, red, green, blue`; *-less* activates `lack`; *green* activates `has, plant`, and the already active `color`; *idea* activates `in, mind, think, make`; *-s* activates `more`; *sleep* activates `rest, conscious`, and the already active `lack`; *furious* activates `angry, er_, gen`. Only *-ly*, a pure category-changing affix, does not activate any element, as it is semantically empty. This is not to say that it entirely lacks a categorial signature: for English, *-ly* is clearly [AN]\D, but in `4lang` we wish to avoid the claim that operators turning adjectives or nouns into adverbs are universal.

In one step we have already activated 20 elements (2.86 per morpheme), and only four of these, `er_, gen`, and `lack` are primitives that resist further spreading, while `in` will invoke the entire `place` conceptual scheme (3.1). In fact, the morpheme count is somewhat arbitrary, as we should clearly add a nominative and an accusative marker, 3rd person singular, present, and perhaps other unmarked operators such as *I declare to you*. To limit combinatorial explosion we need to constrain spreading activation in various ways. First, it is clear that permitting activation in the other direction would be unwise, since one in seven words involve spatial `in`, almost one in three involve possessive `has`, some 40 involve comparative `er_`, some 60 involve negative `lack`, and the same number involve generic `gen`. Second, we need to enforce some condition of locality, in that it is cognitively implausible that the negative element explicit in *colorless* could reinforce the negative element implicit in *sleep* `rest, lack conscious`. We will return to spreading activation in 7.4, where we discuss how to implement locality by island parsing, but we note in advance that the key building block will be the *construction* in the sense of Berkeley Construction Grammar.

`sleep`

For the synthetic compounds in *-er* this superficially takes the form `(N V -er)`$_N$, e.g. in *navel.gaze.er*. Remarkably, the spreading analysis often leads through the unattested intermediary that we use use for agentive *-er*, `(V -er)`$_N$. Once this pattern is activated, the very frequent noun-noun compounding pattern `(N N)`$_N$ can be spread to. In S19:6.4 we wrote

> The algebraic approach (...) largely leaves open the actual contents of the lexicon. Consider the semantics of noun–noun compounds. As Kiparsky (1982) notes, *ropeladder* is 'ladder *made of* rope', *manslaughter* is 'slaughter *under-*

gone by man', and *testtube* is 'tube *used for* test', so the overall semantics can only specify that N_1N_2 is 'N_2 that is V-ed by N_1', i.e. the decomposition is subdirect (yields a superset of the target) rather than direct, as it would be in a fully compositional generative system.

This applies to entries like *teetotaler* which we analyze with an unattested agent noun *totaler* who totals (does always) the V-ing of *tee* (tea). Unsurprisingly (though not exactly predictably) the verb in question is *drink*, so we obtain 'one who always drinks tea'. While still a bit off the actual target 'one who abstains from drinking alcohol', this is close enough for memorization, and offers considerable economy relative to memorizing the entire definition.