STRING DIAGRAMS FOR TEXT

VINCENT WANG-MAŚCIANICA

ST. CATHERINE'S COLLEGE
THE UNIVERSITY OF OXFORD
DEPARTMENT OF COMPUTER SCIENCE

A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
2023

Contents

(Acknowledgements will go in a margin note here.)

1
Continuous relations for semantics

We want to reason formally with and about pictorial iconic representations, of the sort one might draw to solve a problem in elementary geometry stated in words, involving topological concepts such as `touching` and `inside`. To do this in string diagrams, I introduce and investigate the category of continuous relations, **ContRel**.

1.1 Composition of dynamic verbs via temporal anaphora

Dynamic verbs in iconic semantics may be modelled by homotopies, but non-parallel composition of homotopies is only defined up to parameters with indications of how the two separate homotopies begin and end relative to one another; i.e. temporal data.

Example 1.1.1 (Gluing homotopies sequentially at a time $\gamma \in (0,1)$)**.**

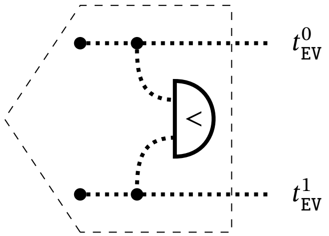
The technical difficulty I'd like to sketch a solution for is that while these parameters must be given as real numbers in the interval $[0,1]$, temporal natural language underspecifies: e.g. in the utterance `Bob drank, and then he slept` he could have drank in the morning and then slept in the afternoon, or both in the evening, and so on. The easy solution is to have absolute temporal anchors, but we seem to get by with less, which appears to necessitate a possible-worlds approach. Arguably the theoretical minimum we require is a kind of algebra for temporal aspects as in Yucatan [CITE], so here I sketch an algebra for temporal anaphora in **ContRel** that only requires copy-delete along with the standard topology on $\mathbb{R}$ obtained by the encoding of intervals as the open set $<: [0,1] \times [0,1]$. Then I'll show how this temporal data can be used to supply the information required for homotopy composition, which should indicate that **ContRel** is in-principle sufficiently expressive for dynamic iconic semantics for natural language, i.e. the interpretation of text as little moving cartoons.

Definition 1.1.2 (A sketch text-circuit algebra for temporal anaphora)**.** We consider three kinds of events.

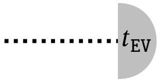
Postscript: with the exception of entification and metaphor, all of these sketches are things I had in mind while initially writing the thesis but didn't make it to the submitted version, so there will probably be technical errors, but this shouldn't compromise the sketches because they are not intended to be rigorous. None of these sketches (and nothing else in this thesis for that matter) should be taken as canonical once-and-for-all solutions to the conceptual problems they are meant to tackle; they are more meant to provoke as first-pass attempts, and they are meant to demonstrate how to play around and have fun in **ContRel** with string diagrams. I'll also note here that everything in **ContRel** is a kind of truth-conditional possible worlds semantics (up to some arbitrary but fixed choice of what particular ensembles of shapes and movements the modeller supplies up front), so there are no guarantees about how any of this material would fare if one tried to take the diagrams and interpret them in terms of neural networks, and I make no claims about whether the mathematics reflects actual cognition. However, I will claim that these mathematical sketches reflect at least the phenomenology of how *I* think about language, which should come as no surprise because my methodology was armchair introspection.

The first is episodic, which corresponds to some interval on $[0,1]$ with endpoints $t^0_{\texttt{EV}}$ and $t^1_{\texttt{EV}}$. We model these as bipartite states with the initial constraint that $t^0_{\texttt{EV}} < t^1_{\texttt{EV}}$. The second is habitual, which could in principle be an arbitrary subset of $[0,1]$, but there are pathologies we would like to rule out as a matter of common sense (e.g. we don't really talk about events that occur in time according cantor set), so we treat habituals as open sets (unions of intervals) to be later constructed or supplied as constraints; when we are finished specifying the algebra, equipping it with unions as a kind of formal sum will approximate those open sets that are constructible by finite amounts of talking about times. The third is a hybrid of the first two, where we consider some open set with distinguished endpoints, modelled as a restriction/intersection of an interval with some other open set.

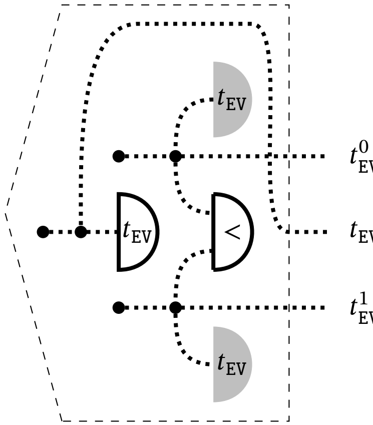
Episodic event
(Interval determined by ordered endpoints)



Habitual event (as constraint)
(An arbitrary open set on $[0,1]$)

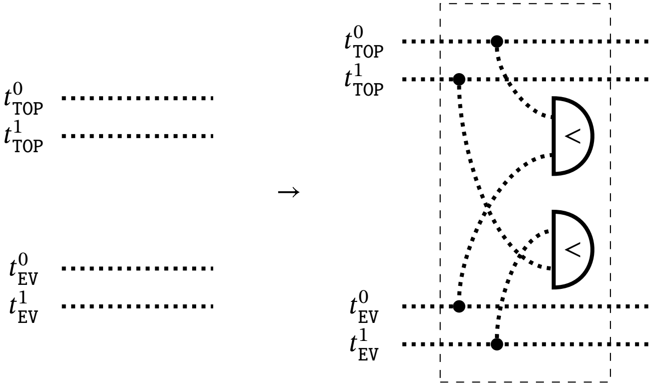


Hybrid event
(Open set with endpoints)



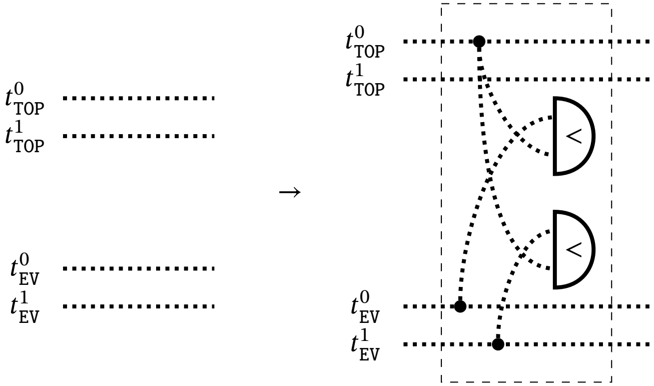
Now we model temporal aspects as circuit components — what appears to distinguish aspects from tenses is that aspects are always relative to the temporal data of two events, whereas tenses may be "intransitive" on events — so all of our aspectual data will involve constraining pairs of events (one of which is a `TOPIC`). The first kind of aspect we consider is *perfective*, which constraints an event time to be within topic time; we model this as imposing a constraint that the endpoints of the event must lie within the interval specified by the endpoints of the topic. In discourse, introducing a perfective constraint corresponds to adding a gate.

Perfective: $t_{\texttt{EV}} \subseteq t_{\texttt{TOP}}$
(Event time contained within topic time)



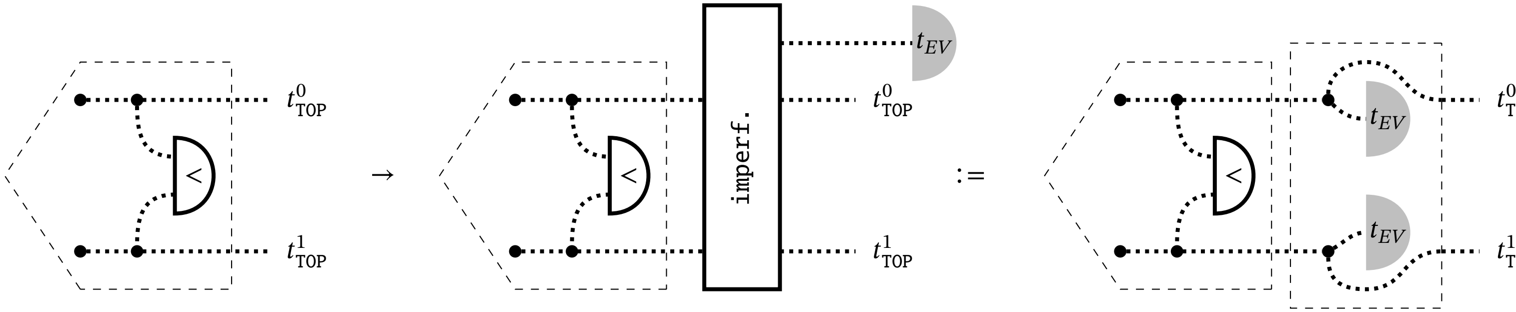
The *terminative* aspect constrains an event to occur entirely before the beginning of the topic time. Terminative composition of verbs may be glossed as `(event) and-then (topic)`, and this kind of composition yields the view of text circuits as implicitly encoding the temporal order in which gate-as-events occur, where now the sequential ordering of gates matters. This failure of interchange interprets text circuits in something like a premonoidal setting [CITE].

Terminative: $t_{\texttt{EV}} < t^0_{\texttt{TOP}}$
(Event will have been completed by the topic time)



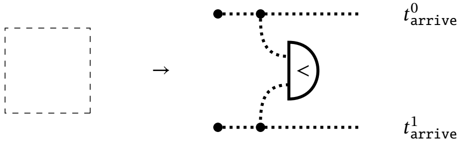
The *imperfective* aspect we consider as constraining an episodic topic time to lie within some ongoing habitual event, where the habitual event is represented as a free coparameter. In discourse, introducing an imperfective constraint corresponds to splicing in such a constraint, which we gloss as a gate that restricts the endpoints of the topic interval to lie within the open set representing the habitual event time as a coparameter. We skip over the subtly distinct *progressive* aspect here as we won't need it for our later example, but it should be clear that an approach along these lines will also suffice.

Imperfective: $t_{\mathtt{TOP}} \subset t_{\mathtt{EV}}$
(Episodic time contained within habitual event time)



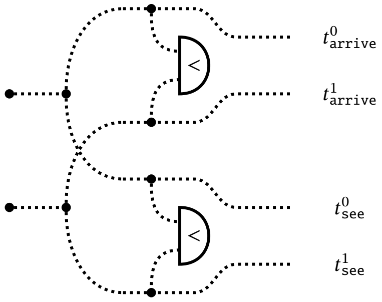
Example 1.1.3. So here is an example of Yucatan Maya taken from [CITE], which is an excerpt of an interview with a speaker fleeing a cyclone. I have split the excerpt into numbered single-verb clauses, accompanied by glosses in English with aspect-markers and the corresponding evolution of a text-circuit by the discourse rewrites we have defined. The first event introduced into discourse is the arrival of the refugees in the village, which is marked as perfective.

(1) $\left[\dfrac{\textit{Kaajk'ucho'on}\ \texttt{túun way tekàajil x Jaxleyile,}}{\texttt{When we}\ \underbrace{\textit{arrived}}_{\text{PERF.}}\texttt{,}}\right]$



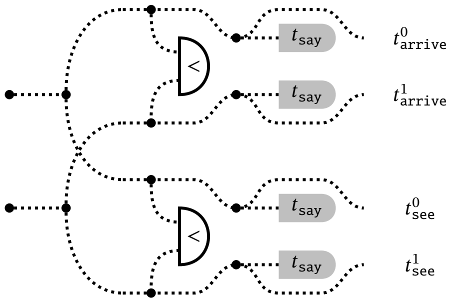
The second event is what the refugees saw, implicitly concurrent with event (1), which we opt to treat with a prepended copy of endpoints. `arrive & see` then form an atomic topic for events (3) and (4), which we deal with by constraining both (1) and (2) in the same way. Note that there is a single variable open set $t_{\mathtt{say}}$ that is repeated 4 times in the diagram.

(2) $$\left[\frac{\textit{kilike'}\ \texttt{tuláakal máake',}}{\texttt{we}\ \underbrace{\textit{saw}}_{\text{PERF.}}\ \texttt{,}}\right] \rightarrow$$



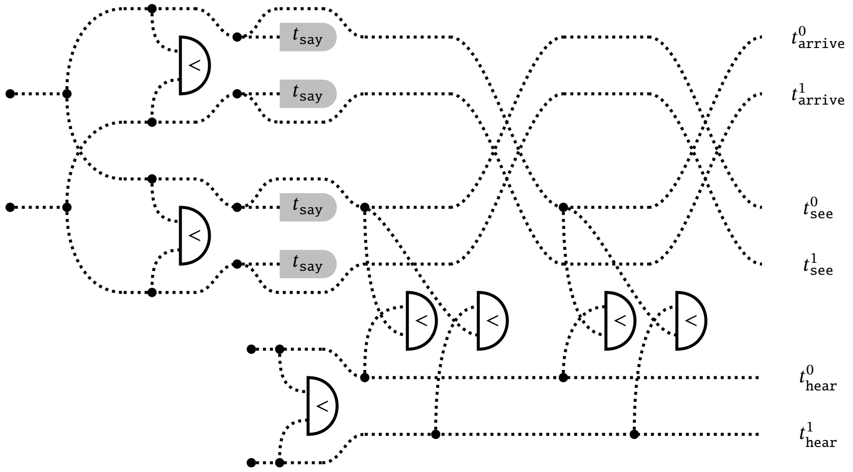
The third event refers to the villagers saying something, in the imperfective aspect with respect to events (1) and (2), so we constrain those topics accordingly. In gloss, it was an ongoing event that the villagers were saying something when the refugees arrived.

(3) $$\left[\frac{\textit{táan uya'aliko'obe'}\ \texttt{jach}}{\texttt{everyone}\ \underbrace{\textit{was saying}}_{\text{IMPF. WRT. (1,2)}}}\right] \rightarrow$$



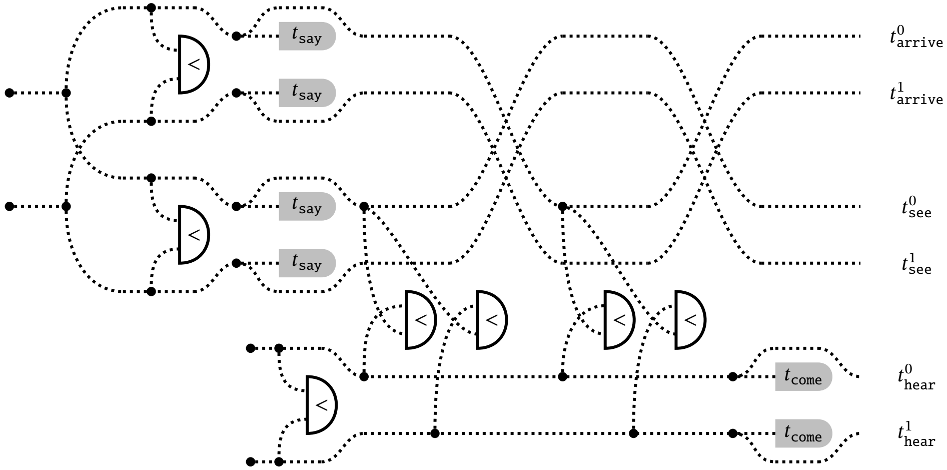
The fourth event refers to what the villagers had heard, in the terminative aspect with respect to (1) and (2). In gloss, the villagers were saying (reporting) the episodic event of them hearing something on the radio, and this hearing-event had completed before the refugees' arrival.

(4) $\left[\begin{array}{c}\texttt{\textit{ts'uyu'ubiko'ob ti'} ràadyoe'} \\ \hline \texttt{(they)}\ \underbrace{\texttt{\textit{had heard}}}_{\text{TERM. WRT. (1,2)}}\ \texttt{on the radio}\end{array}\right]$ →



The fifth event refers the coming of the cyclone, which was ongoing at the time of the villagers hearing the radio report. This introduces a new habitual event as the variable open set $t_{\mathtt{come}}$, repeated twice in the diagram as constraints.

(5) $\left[\begin{array}{c}\texttt{\textit{túun tàal} le siklòono'.} \\ \hline \texttt{the cyclone}\ \underbrace{\texttt{\textit{was coming.}}}_{\text{(4) IMPF. WRT. (5)}}\end{array}\right]$ →



Altogether, the final diagram represents a map from two open sets on $[0,1]$ (representing the potentially habitual events `say` and `come` encoded as variable open sets $t_{\mathtt{say}}$ and $t_{\mathtt{come}}$) to return a state in **ContRel** that encodes the set of possible endpoints for the episodic events `arrive`, `see` and `hear`: $\{(t^0_{\mathtt{arrive}}, t^1_{\mathtt{arrive}}, t^0_{\mathtt{see}}, t^1_{\mathtt{see}}, t^0_{\mathtt{hear}}, t^1_{\mathtt{hear}})\}$. Moreover, we have set up the algebra to allow us to leverage compositional discourse structure in such a way that sampling any of the elements of the resultant set returns a choice of endpoints consistent with the temporal constraints of the excerpt.

1.2 Iconic semantics for modal verbs

In this sketch I want to deal with certain modal verbs: that means those of cognition and perception like to think and see, and the sketch will taper out towards some modal auxiliaries like wanting. These kinds of verbs are roughly characterised as requiring copies of entities to be instantiated in worlds similar to but not exactly that of whatever base narrative reality is referred to in the discourse. For example, in `Alice sees Bob drink a beer, Bob drinks another after Alice leaves.`, there are two Bobs, because the one in Alice's mental-theatre drinks a single beer, and the one in the base reality of the narration drinks two. So there are two worlds $\mathfrak{W}$ here, one basic, and a $\mathfrak{W}_{\mathtt{A}}$ for the world in Alice's perception. Things get intractably tricky fairly quickly with these modals: to do epistemic logic means to have nested indices of what Alice thinks Bob thinks Alice thinks, to gossip is to reason about he-said-she-said, to understand complex narratives is to reason about stories-told-within-stories, and counterfactuals are a whole thing too. So that is a fundamental mystery: all this seems fairly complicated to encode and reason about symbolically, but it is phenomenologically fairly easy for adults to do, so what gives? What sort of mathematical presentation of these modals would at least reflect this lightness and ease?

I think thought-bubbles that show up in comic books are a pretty good start. Their cloudlike shape is a visual convention indicating a separate mental world, and they are typically used to represent want when the contents are also iconic representations.

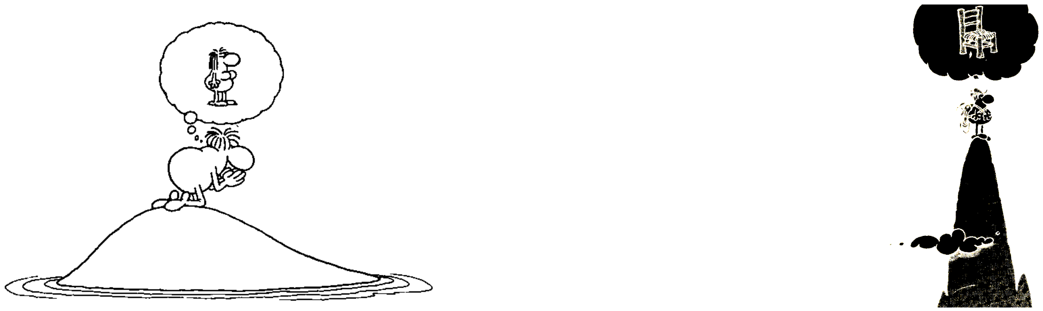


Figure 1.1: Two examples by Mordillo, an artist I liked as a child: a thought bubble representing a woman, where the context of a stranded man implies a want for companionship, and a thought bubble representing a chair, where the context of a climber on a tall summit implies a want for rest.

The visual convention for cognitive and perceptive-alethic verbs is, as far as I can tell, a kind of x-ray effect into the contents of a head, which employs the familiar container metaphor: the head is a container for thoughts.

For alethic verbs in particular (those modals that are truth-preserving, in that they "do not forget" the truth), there's a need for the contents of the container to be synchronised with the contents of the outside world. Here are some observations that enable this in **Contrel**. The basic enabling insight is that, in Euclidean spaces, if we have a hollow container with a solid blob inside, there's an approximately continuous bijection between the (open set) insides of the container and the outside world.

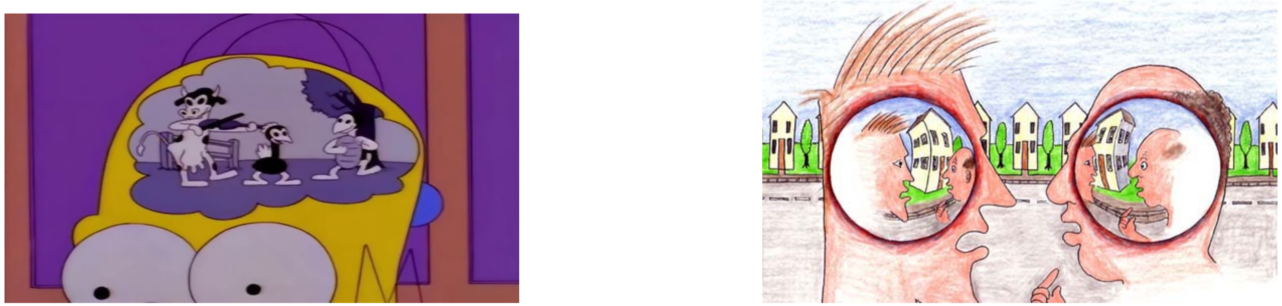


Figure 1.2: On the left, a scene from the Simpsons showing the contents of Homer's mental-theatre. On the right, a depiction of two separate mental-theatres with a fisheye effect, taken from Steven Lahars "A Cartoon Epistemology" freely available online, which was also the initial inspiration for this sketch.

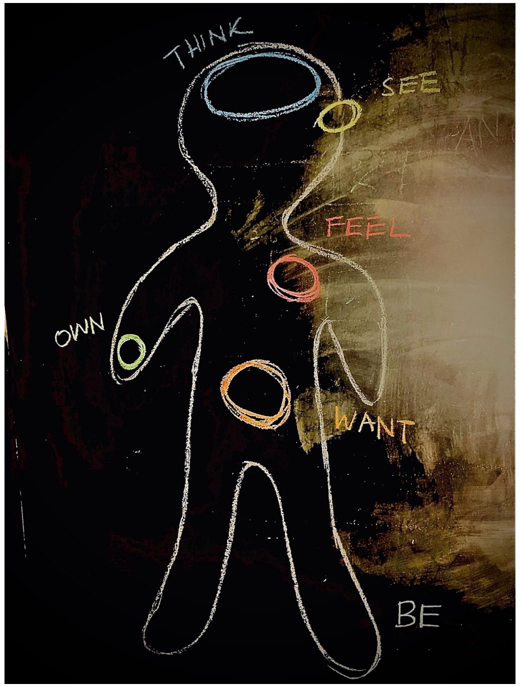


Figure 1.3: So the basic idea is to put representations of worlds inside bounded regions as containers, and in this way iconic semantics provides a univocal setting that displays all of the relevant worlds at once. We are free to pick visual conventions, as they are no more or less arbitrary than the assignment of indices and symbols such as $\mathfrak{W}_A$ to the contents of possible worlds. Here is a sketch convention for containers on an iconic representation of a person for different modal verbs: seeing, thinking, feeling, owning, and wanting. I sent this excitedly with little supporting context to Bob while I was writing my thesis. He was concerned. Then I got concerned. Childlike became creepy, and neither are good looks. I think I have supplied enough context to make this sensible, but there's no way I'm going to beat the crazy allegations.

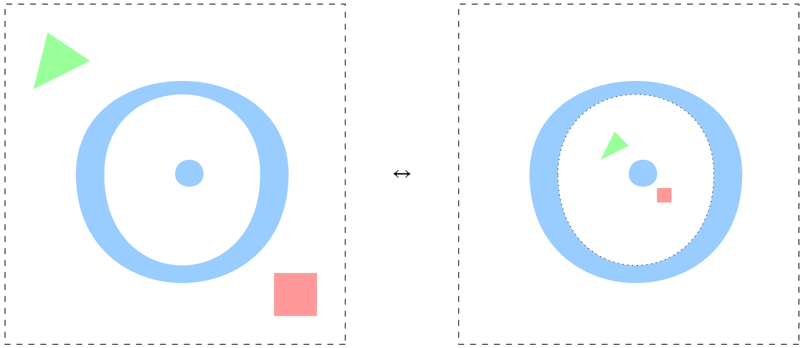


Figure 1.4: The inside and the outside of a container with a solid blob inside are both homotopic to the space with a puncture. This is only approximately a continuous bijection because the unbounded outside space can only map to the open interior of the container. We can use such bijections as a bridge to establish connections between elements of different possible worlds.

The second, and unfinished, idea is that if we have a handle on the individual components of sticky spiders, then we may use something like a very-well-behaved lens (hence its occurrence in the introduction) to ensure that the inside of the container is really behaving like a faithful storage medium for the goings-on outside. I think that's suggestive enough, and I'll deal with parthood in the next sketch. The last thing I want to deal with here is the problem of infinite regress for epistemic modals like knowing: if I know something, then I know I know it, and I know I know I know it, and so on. A naïve solution is to just use an infinitely-nested series of containers.

Figure 1.5: Again from Cartoon Epistemology, on the unsatisfactory nature of infinitely-nested containers: *But who is the viewer of this internal theatre of the mind? For whose benefit is this internal performance produced? Is it the little man at the center who sees this scene? But then how does HE see? Is there yet another smaller man inside that little man's head, and so on to an infinite regress of observers within observers?*

So the problem here is how to encode this infinite regress with finite means in an iconic model. The usual monadic approach still runs into the problem that you have to map a potential nested-infinity of possible worlds onto some finite model if one cares about cognitive realism. In iconic semantics, we can modify the space itself; here I think Escher was onto something.

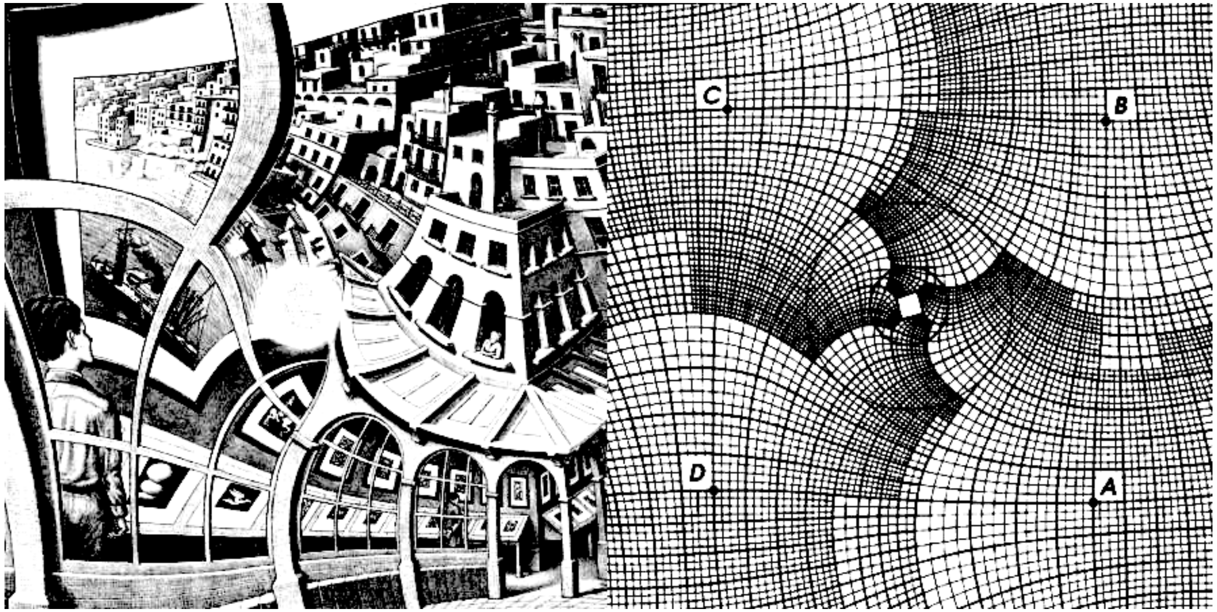


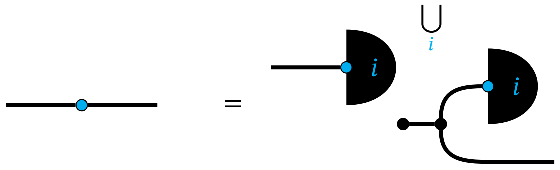
Figure 1.6: Escher's "Print Gallery" lithograph alongside his working sketch of the vortex-grid geometry the work was built on. On the left of the lithograph, an observer examines a framed painting of a town. Going clockwise, we see more details of the town, which has in it a print gallery, within which is the original observer. The missing centre of the piece where Escher signed the work obscures what would have been infinite nesting; the right-hand-side of the frame would have spiraled along the vortex infinitely. Treating the frame as a container, here we have an example of a container that contains itself, where movement clockwise indicates going down a level, clockwise going up, yet no explicit infinities anywhere.

The space in which such an arrangement can be realised is the same as that of the Penrose staircase: splitting the lithograph into four corners, each is a locally consistent snapshot, each gluing of quadrants is a consistent (as/de)scent, but the overall manifold obtained needs to be embedded in a higher dimension. While this in principle solves the problem of finitely representing infinite descent, these kinds of spaces are not grounded in physical, embodied intuitions. I think it is mathematically neat that there can exist topological models for such modal verbs, but whether such proposals are to be taken seriously as modelling cognition is a thorny matter I don't want to say more about.

1.3 Configuration spaces

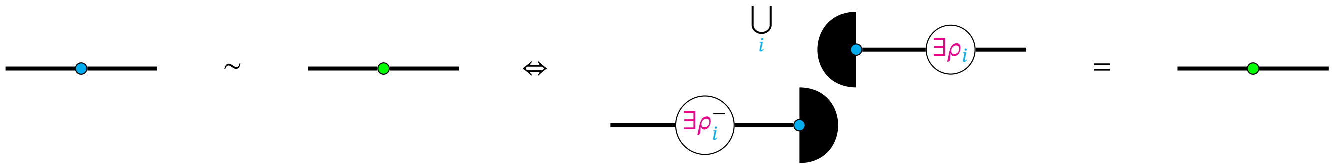
Individual sticky-spiders correspond to static collections of set-labelled shapes in **ContRel**; in this sketch I want to talk about all the different ways the same collection of shapes can be arranged in space. Let's assume for simplicity that we only want to deal with *nice* sticky-spiders where cores and halos agree and are both contractible opens; i.e. the spider can be expressed as a finite union of open solid blobs as effects followed by the same open solid blob as a state.

Definition 1.3.1 (Nice sticky-spiders)**.** A sticky-spider is *nice* if it is equal to a union of contractible open effects followed by the same contractible open expressed as a state.



Let's also say we start with the ability to detect whether two sticky-spiders are related to one another by rigid displacements, expressed as a topological group with elements we denote ρ. Since sticky-spiders can be represented as unions of effects followed by states, we can define a binary relation on sticky-spiders that tells us whether they are the same up to rigidly displacing component shapes:

Definition 1.3.2 (Displacement relation)**.** Two sticky-spiders (cyan and green, both assumed to be nice here), each with components indexed by I, are *equivalent up to displacement* when there exist ρ_i such that:



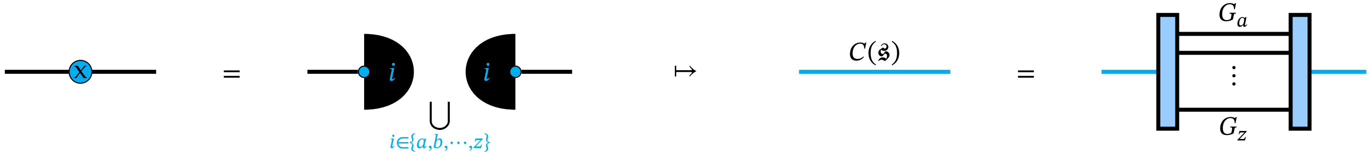
We've suppressed labelling of the states and we've contracted the cup to just depict the open state as a semicircle.

Displacement is evidently an equivalence relation, and moreover requires that the two spiders related have the same number of components. Now given a particular nice spider, we treat its equivalence class of spiders as a configuration space in which we have access to all of its rigidly displaced variants at once.

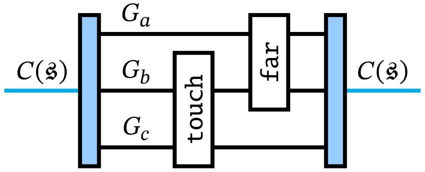
Definition 1.3.3. The *configuration space* $C(\mathfrak{s})$ of a nice spider $\mathfrak{s}$ with indexing set I is the topological space with underlying set defined to be the equivalence class $[\mathfrak{s}]$ of $\mathfrak{s}$ under displacement. Assuming the topological group of rigid displacements is itself a topological space G, the topology of $C(\mathfrak{s})$ is a restriction of $\bigtimes^{|I|} G$

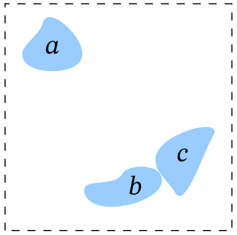
to those $|I|$-tuples of displacements witnessed by $[\mathfrak{s}]$.

In configuration spaces we're making use of the fact that any displacement relationship comes with (up to a non-unique choice of basepoints for each component shape) a witnessing tuple of ρ_is. As a consequence, the configuration space of a sticky-spider is a retract of the product space $\bigtimes^{|I|} G$ where G is the topological group of displacements, and we can use the identity relation between the section and retraction to strip the configuration space wire, revealing each of the $\bigtimes^{|I|} G$ like guitar strings: each element of the set that the initial nice spider $\mathfrak{s}$ splits through gets its own string.

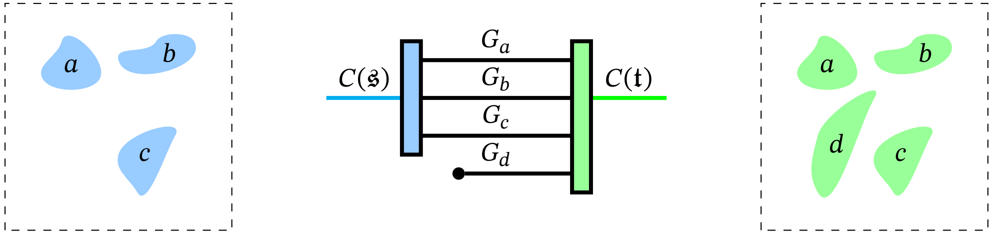


Note that although every guitar string is G, there is extra typing data indicating which element of the indexing set of the spider each G corresponds to. So here's a model in which the named wires of text circuits make sense. We can put gates on the guitar strings, which may for example correspond to constraints on the relative positions of shapes in configuration space.





The next thing we can try is to add and subtract shapes from configuration spaces, and while there are technical details like matching choices of basepoints I'll gloss over, the gist is this: when the shapes in a nice spider $\mathfrak{s}$ are a subset of the shapes in a nice spider $\mathfrak{t}$, we can add in states to the guitar-picture of $\mathfrak{s}$ and wrap them up again using the idempotent of $\mathfrak{t}$, and we can delete wires in the guitar-picture of $\mathfrak{t}$ and wrap that up using the idempotent of $\mathfrak{s}$.



The last stop in this sketch is disintegrating and integrating shapes; if we could freely break apart a shape, we know that in principle we get another configuration space where we can manipulate those parts, and if we can glue those pieces back together again, then we could do simple things like open and close containers. Let's first define the disintegration relation between spiders. Observe that the data of a nice spider is equivalently viewed as a function $f : I \to \mathfrak{O}$, where I is the indexing set, and $\mathfrak{O}$ is some set of opens with whatever well-behaviour condition, along with the constraint that $f(x) \cap f(y) \neq \varnothing \Rightarrow x = y$ that enforces non-overlapping shapes. This perspective gives us a foothold to define a disintegration relation: a "more refined" spider is one that has a superset of I as domain, with a function that sends elements of the indexing set to either the same shape as f, or a subshape.

Definition 1.3.4 (Disintegration)**.** Let $\mathfrak{s}$ and $\mathfrak{t}$ be nice spiders, described by functions $s : I \to \mathfrak{O}$ and $t : J \to \mathfrak{O}$ respectively. $\mathfrak{t}$ *disintegrates* $\mathfrak{s}$ if there exists a surjective $d : J \twoheadrightarrow I$

2

Bibliography