Artificial Understanding

MSc Thesis *(Afstudeerscriptie)*

written by

Arie W. Soeteman
(born July 4th, 1996 in Aerdenhout, Nederland)

under the supervision of **prof. dr. Michiel van Lambalgen (ILLC, supervisor)**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:
July 6th, 2022

Members of the Thesis Committee:
prof. dr. Michiel van Lambalgen (ILLC))
dr. Nick Bezhanishvili (ILLC)
dr. Ronald de Haan (ILLC)
dr. Richard Evans (Google DeepMind London, external expert)
dr. Malvin Gattinger (ILLC, chair)

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Abstract

This thesis explores the processes an artificial agent needs to understand its environment. It extends on the Apperception Engine and intensively applies insights from Kant's *Critique of Pure Reason*. Techniques from logic programming, topology, graph theory and several other disciplines are harnessed to bring these two frameworks into further correspondence with one another. The result consists of two computational systems. The first is a direct extension of the Apperception Engine that uses *geometric logic* to express Kant's functions of judgement. The second is a Figurative Apperception Engine that implements Kant's *spatio-temporal* or *figurative synthesis*: input is taken up and combined in a unifying process that builds both space and time as qualitative structures. By applying program synthesis within a Kantian architecture, a step is made towards the development of artificial agents that are both explainable and generally competent.

Acknowledgements

I must express my deepest thanks to Michiel. Your patience and open mind, incomparable knowledge and never faltering ambition have shaped this thesis from beginning to end. I am also very grateful to Richard, who has taught me not to be afraid of big questions, and has so often identified the weakest spots with extreme precision. I would like to express my gratitude to the thesis committee. Plowing through this thesis in full is not a small task, and feedback from educated minds is a very valuable commodity for a starting researcher. My parents cannot be forgotten, who have helped me find the rest to develop new ideas, and have provided many dinners, which we ate together while I stared blankly into space thinking about mathematics. Finally, I thank my dear Anne. Your support has always been with me to help me persevere, and your presence has helped me to enjoy life, even if there was very little time to do so.

Preface: Understanding Artificial Agents

When does an agent understand its environment?

This question, that has been a central topic from Greek epistemology to modern Neuroscience and AI, has received a wide range of different answers. Leaving the term 'agent' now vaguely defined as *something with the ability to act*, the answer primarily hinges on your desired definition of understanding, and from it flow debates on subjects such as agency, responsibility and consciousness. The Cartesian *res cogitans* has long been a central view in which understanding fills its own realm, separated by an unbridgeable gap from material reality, and only accessible to itself. This Cartesian division constituted a firm ground for our freedom and autonomy as understanding agents in a world that is governed by the deterministic laws of physics, and simultaneously restricted these remarkable qualities to ourselves, barring non-human agents from the spectrum of understanding.

An almost opposite view on understanding is adhered to by some modern AI designers, aiming to build agents with this quality from logical circuits and statistical systems. Shane Legg and Marcus Hutter for instance, proposed in their influential article: "Universal intelligence" (2007) that intelligence is the measure of "an agent's ability to achieve goals in a wide range of environments" (p.12). From this perspective, in which intelligence is the measurable result of capacities such as planning and reasoning, the question of whether or not this result is accompanied by an internal understanding seems to become an irrelevant complication:

"From our perspective, whether or not an agent understands what it is doing is only important to the extent that it affects the measurable performance of the agent ... indeed it is not even clear to us how to define "understanding" if its presence has no measurable effects". (p.49).

Unfortunately, if one merely takes understanding to be an elusive source of behavior, it seems to be a difficult concept to activate in a useful manner. Its common association with subjective experience makes it a favored topic for AI critics under the flag of Searles Chinese room: *We* clearly understand what we're doing, but artificial agents *just do*. At the same time, this subjective association strips understanding from any objectively measurable characteristic, inducing designers to drop the subject altogether and simply build agents that work. The question of who understands what and why is then reduced to a classical yes-no debate, preventing any real progress on the topic.

Activating understanding

An interesting alternative for this view that will be taken as a central motivation for this thesis is provided by Daniel Dennett, who claims that understanding is "not the source of competence or the active ingredient in competence" (2018). Instead, he argues, understanding is *composed* of competences. The idea that our actions have an overseeing understanding as their source is merely a consequence of the idealised view of ourselves and others as intentional agents, i.e. as having reasons and arguments for what we do. While this *intentional perspective* can be a useful tool for explaining and predicting behavior, it is also in many cases blatantly misleading. How often do we act competently, for instance when calling a bluff in poker or buying a profiting stock, without knowing exactly why we did such a thing? And how easily do we explain such actions afterwards by an intelligent argument showing that we 'understood' what we were doing all along? Because we, according to Dennett, so often mistakenly take the successful intentional perspective to be a truthful explanation of how our actions came to be, we come to believe that understanding is a modus of subjective experience, and thus only available to things that have *consciousness*. Since consciousness is, in turn, often considered to be an all-or-nothing quality, the same unfortunately follows for understanding.

Dennett rejects this idea of understanding as a subjective experience that brings forth competence, and proposes a way out of the deadlock. If understanding is composed of competences then we can, like competence, encounter it in endlessly many configurations and gradations. All we have to do is provide adequate descriptions to categorize this field. He provides one simple proposal of such a description using four distinct classes. *Darwinian creatures* are gifted with a fixed set of competences, provided by some external source such as natural selection or intelligent design. While these creatures might be extremely competent, they do not seem to understand anything by any commonsense definition of the term, as can for instance be shown by slightly changing their environment, which renders them helpless. *Skinnerian creatures* additionally have the possibility to adjust their behavior based on positive and negative stimuli. They display some aspect of non-regularity in their behavior so that positively rewarded actions are stimulated. Although these creatures have been supplied with the means to adjust and adapt themselves to some extent, they are generally not considered to understand why they do this at all. *Popperian creatures* have some way of storing information about their environment, so that they may simulate their actions before executing them. As Popper expressed this: they let their hypotheses die instead of themselves. Such creatures are often considered to have at least some form of understanding. Finally, the *Gregorian creatures* are described by Dennett as using "tools for thinking", which can be theories and words as well as calculators and telephones. They have the ability to actively select those tools and solutions that are best suited to the problems that they encounter, and to decide which tools are worth mental effort such as memory storage or communication.

Without further exploring here the interesting questions of which agents belong to which class and how we can more precisely distinguish their competences, I pause here to note that this view on understanding seems to constitute a more fruitful scientific perspective then viewing it as a capacity that some simply do, and others simply do not have. Indeed, I believe that characterising understanding as *composed of competence*, directly provides more structure to the definition of intelligence provided by Legg and Hutter. The agents we know that display the competences that satisfy their criterium of "achieving goals in a wide range of environments" seem to be precisely those agents of which we generally believe that they can bring forth some level of understanding. An essential aspect of Legg and Hutter's definition is the use of the word *wide*, which emphasizes generalizability as a core aspect of intelligent behavior. The actively established qualities displayed by Dennett's Popperian and Gregorian creatures such as model constitution, simulation and active selection of solutions all seem to represent essential (or at least useful) subsets of the totality of competences that comprise, from a wider perspective, the activities that Legg & Hutter call 'intelligent'. **We shift here from directly evaluating an agent's behavior to evaluating which active processes an agent needs to implement in order to constitute a mode of operation (read: understanding) that brings forth the desired behavior**.

As will be made clear in the upcoming chapters, analysis of the processes required to bring forth understanding has been a central topic in Kant's *Critique of Pure Reason*, and likewise it is a central theme behind this thesis.

Artificial Agents

Now that we have a more thorough understanding of understanding, we can apply this to evaluate artificial agents. Where do these systems stand on Dennett's scale? Traditionally, a crude distinction is made in AI between *symbolic systems* (good old-fashioned AI or GOFAI), and *machine learning systems*. If we momentarily forget about the many examples of impressive improvements on and combinations of these two types of system, we can make a caricature of both which might teach us more about their 'artificial understanding'. Symbolic systems then make use of formal descriptions, logical inferences and predefined rules. Typically, designers, with an extensive understanding of the task that their systems are meant to perform, supply them with a predefined and unchanging set of sensors and rule structures that allow them to act competently within their domain. These good old-fashioned systems can thus be characterized as Darwinian creatures: a clear example of competence without understanding. They have no ability to adapt, and small changes in their environment may render them hopelessly incompetent.

Machine learning systems, on the other hand, make use of statistical learning to capture regularities in data. Their designers provide them with optimization functions and expose them to sufficiently many stimuli so that they may optimize their parameters and consequently their behavior. These systems thus seem to fall in the category of Skinnerian creatures: they can adapt, more so than GOFAI, but they do so by directly applying the stimulus-responses that were provided to them beforehand. It would be terribly difficult to argue that they formulate their own hypotheses or simulate their environment. We can now stop forgetting about the crudeness of the made distinction, and nonetheless note how both system types have complementary strengths and weaknesses. Symbolic systems are easy to interpret. They typically do not require large datasets to train and can reason easily about general concepts, as long as these concepts have been properly supplied to them by their designers. Machine learning systems on the other hand are notoriously difficult to interpret and explain. They require extensive and often expensive training, and are very difficult to apply to general problems. However, they are very good at working with noisy input, a trait that symbolic systems lack.

Of course many novel systems have been developed that aim to combine the benefits of both approaches. Providing agents with the ability to learn both statistical correlations and general conceptual reasoning is seen as one of the most promising paths towards the next breakthrough in artificial competence (e.g. (Marcus, 2018)). Supervised learning systems can be supplied with smartly chosen (symbolic) labels and domain-specific learning architectures, so that they may learn faster and better, and apply their Skinnerian optimized parameters within the context of a competent Darwinian structure. However, one may doubt whether such approaches bring artificial agents further on the scale of understanding, regardless of the impressiveness of their competences and the understanding thereby displayed by their designers. It is a whole different matter to provide agents with the capacity to actively formulate and reject their *own hypotheses*, and to perform their *own generalisations*. If we adopt the hypothesis that Dennett's scale of understanding specifies the modi operandi that come with acting 'in a wide range of environments', it might prove to be very difficult to achieve this goal without taking understanding seriously. This thesis explores the processes needed for *understanding trough model construction*, extending on a pioneering AI system, and applying conceptual frameworks developed by Kant in his critique of metaphysics. I hope the reader will enjoy the insights developed in this work, and I excuse its lengthy nature using the same words of Jean Terasson that were harnessed by Kant:

"if the size of a book is measured not by the number of pages but by the time needed to understand it, then it can be said of many a book that it would be much shorter if it were not so short." (Translation, (Kant, 1998))

Introduction

This thesis builds upon two main sources of inspiration. The first is the APPERCEPTION ENGINE (AE): a system that understands its environment through logical model construction (Evans, 2020). The second is Kant's Critique of Pure reason (CPR): the redirection and limitation of methaphysics that was written by Kant until 1787. The APPERCEPTION ENGINE draws on the *cognitive architecture* expounded in the Critique, for its structure and *domain general inductive biases*. The aim of this thesis is to bring these two frameworks even further together, translating more of Kant's heritage into mathematical formalisms and computational systems.

The first chapter gives a general background of logic programming, the APPERCEPTION ENGINE and the *Critique*. The second chapter compares the two main pillars that were just introduced in a more intensive manner. The result is an enumeration of difference between the AE and CPR. The rest of this thesis then aims to resolve a subset of these differences. The third chapter addresses the structure of *judgement*, implementing *geometric logic* as a more expressive language for model construction that fits the Kantian framework. The fourth, fifth and sixth chapters aim at representing synthesis in space and time. The capacity of an agent to understand what it is given in a spatio-temporal manner is then argued to underlie the constitution of experiential unity. Throughout this thesis, all philosophical and mathematical arguments are developed with the aim of computational implementation in mind, so that many theoretically interesting cases such as continuous inputs and infinite models are often not taken into consideration without any explicit justification. The less experienced reader will hopefully find that all significant concepts have been properly defined and introduced either in the first chapter, or where they are relevant. The more critical reader will hopefully be content with the overview of bounds and limitations provided in the discussion. The reader or grader can find the source code for chapters 4, 5 and 6 at https://github.com/ariesoeteman/figurative_apperception_engine. The code is however not yet properly structured and documented, so it is merely intended to provide a general overview of what has been done. A proper application will be posted online in the near future. Now let's begin.

Contents

Chapter 1

Background

1.1 Logic programming

Answer Set Programming (ASP) is a logic programming language based on the stable model semantics. Its programs contain:

- **Predicate symbols** and **function symbols** with an associated arity, denoted by strings starting with a lowercase letter (p, f, even, m_OTHER).
- **Constants**, also denoted by strings starting with a lowercase letter (a, john, number3).
- **Variables**, denoted by strings starting with an uppercase letter (X, Node, ATOM).

In this thesis I usually denote, unless context demands otherwise, predicate symbols with $p, q, r \ldots$, function symbols with $f, g, h \ldots$, constants with $a, b, c, \ldots$ and variables with $X, Y, Z \ldots$ in ASP and x, y, z in mathematical reasoning.

A *term* is a constant or variable or of the form $f(t_1, \ldots t_n)$ where f is a function and the t_i are terms. An *atom* is of the form $p(t_1, \ldots, t_n)$ where p is a predicate and the t_i are again terms. We call variable-free terms, atoms and programs *ground*.

A *normal logic program* over a set of ground atoms is a finite set of rules of the form:

$$a_0 : -a_1, \ldots, a_m, \text{not } a_{m+1}, \ldots \text{not } a_n.$$

Where each a_i is a ground atom. a_0 is called the head, and $\{a_1, \ldots a_m, \text{not } a_{m+1}, \text{not } a_n\}$ is called the body of the rule. These are also denoted as head(r) and body(r). A rule of this form expresses that if all literals in the body are true the head must also be true, where not a_i is true if a_i has not been proven to be true. Thus, 'not' expresses negation by default. If the head of a rule is empty it is taken as $\bot$, and we call the rule a *constraint*. If the body of a rule is empty it is taken as $\top$, and we call the rule a *fact*. Associating each rule r with sets $body(r)^+ = \{a_1 \ldots a_m\}$ and $body(r)^- = \{a_{m+1}, \ldots a_n\}$, a set of ground atoms X satisfies the body of a rule r in P ($X \models body(r)$ if $body(r)^+ \subseteq X$ and $body(r)^- \cap X = \emptyset$. If $X \models body(r)$ implies $head(r) \in X$, we then say that X satisfies r, and if X satisfies all rules in P we call it a *model* for P.

Given a (generally unground) normal logic program P, its *Herbrand universe* is the set of all ground terms that can be formed from the constants and function symbols in P. Rules in unground programs must be *safe*, meaning that the variables in the head of a rule are a subset of the variables in its body. The ground instance $grd(r)$ of a rule in P is the set of all ground rules obtained by substituting its variables with ground terms in the Herbrand Universe. Ground programs $grd(P)$ are defined accordingly. The *Herbrand Base* is the set of all atoms constructed from the predicates in P and the terms in its Herbrand Universe. A *Herbrand Model* for P is then a subset of the Herbrand base that is a model of $grd(P)$.

Given a ground normal logic program P and a set of ground atoms M we can define the single-step consequence operator:

$$T_P(M) = M \cup \{head(r) : r \in P, M \models body(r)\}$$

A *positive logic program* only contains rules where $body(r)^- = \emptyset$. It can then be shown that each positive logic program P has a unique $\subseteq$ minimal Herbrand model as the least fixed point $T^{\infty}_{grd(P)}(\emptyset)$.

Answer Set Programming makes use of the stable-model semantics. Given a ground normal logic program P and a set of ground atoms X we define the reduct of P as:

$$P^X = \{head(r) :- body(r)^+ \; : \; r \in P, body(r)^- \cap X = \emptyset\}$$

Intuitively, we remove all rules that contain a literal 'not a_i' in the body such that $a_i \in X$, and then remove all literals 'not a_i' from the bodies of the remaining rules. A stable model, or *answer set*, for a ground normal logic program P is then a set of ground atoms X such that $T^{\infty}_{P^X}(\emptyset) = X$. Accordingly, if P is unground its stable model satisfies $T^{\infty}_{grd(P)^X}(\emptyset) = X$. Normal logic programs can have many stable models. Each stable model is however $\subseteq$ minimal, in the sense that no stable of the same program can be a proper subset.

In *Kant's Cognitive Architecture* (2020), ASP is used as an interpreter or meta language for the object language DATALOG$^{\ni}$, a simple extension of DATALOG that allows for constraints and causal implications. Rules in DATALOG are denoted as:

$$a_0 \leftarrow a_1, \ldots, a_m.$$

Where the a_i are *function-free* atoms. DATALOG programs thus correspond to function free positive logic programs, and have a unique minimal model corresponding to a unique total model (where the negation of untrue atoms is derived explicitly), that can be constructed by the consequence operator T.

ASP allows for several additional syntactical structures such as *choice rules* and *optimization statements*. Choice rules are of the form:

$$k\,\{a_1, \ldots a_m\}\,l \; :- \, a_{m+1}, \; \ldots, \; a_n, \text{ not } a_{n+1}, \ldots \text{not } a_o.$$

Where if the body of the rule is satisfied, now any subset S of the atoms in the head such that $k \leq |S| \leq l$ can be derived. Sets of atoms with an upper or lower bound can also be placed in the body of a rule, where they are satisfied if and only if the number of true atoms in the set is between the specified bounds.

Optimization statements are of the form

$$maximize\{l_1 = w_1@p_1, \dots w_n@p_n\}$$

Here the w_i represent the weight assigned to each associated ground instance of l_i, and p_i represents a priority where higher priorities are maximized first. For a complete overview of the syntactical possibilities of ASP one can read for example *Answer Set Solving in Practice* (Gebser, Kaminski, Kaufmann, & Schaub, 2012).

The *complexity* of a logic program usually denotes the time-complexity for checking whether $P \models A$ as a function of the size of program P and set of ground atoms A (Dantsin, Eiter, Gottlob, & Voronkov, 1997). The ground instantiation of a logic program is in general exponential in its size, so that the logical consequence problem for DATALOG is EXPTIME complete. For ASP, function symbols can give rise to infinite models, so that model existence is in general undecidable. If function symbols are not allowed, a distinction is made between brave reasoning, where $P \models A$ if P has a model containing A, and cautious reasoning, where $P \models A$ if all models of P contain A. Brave reasoning clearly corresponds to model existence. Allowing unrestricted use of 'not' and disjunctions in the head of function-free ASP programs results in a brave complexity that is $\text{NEXPTIME}^{\text{NP}}$ complete and cautious complexity that is $\text{CO-NEXPTIME}^{\text{NP}}$ complete (Eiter, Leone, & Sacca, 1998) (Eiter, Faber, Fink, & Woltran, 2007).

Often, complexity is analysed with respect to ground programs. In this case, complexity for DATALOG is P complete. The brave and cautious reasoning tasks for ASP without disjunctions are NP and CO-NP complete respectively. If we further allow disjunctions in the heads the reasoning tasks are $\Sigma_2^P = \text{NP}^{\text{NP}}$ and $\Pi_2^P = \text{CO-NP}^{\text{NP}}$ complete, and if we then also include weak optimization statements this gives $\Sigma_3^P = \text{NP}^{\Sigma_2^P}$ and $\Pi_3^P = \text{CO-NP}^{\Sigma_2^P}$ completeness (Eiter et al., 2007). In this thesis, unless explicitly stated otherwise 'complexity' refers to brave reasoning complexity with respect to ground programs.

1.2 The Apperception Engine

I here give a brief overview of the APPERCEPTION ENGINE, a system developped by Richard Evans that uses ASP to construct theories that make sense of a sensory sequence (2020). I briefly explain how Evans defines 'sensory sequence', 'theory' and 'making sense'.

The *sensory sequences* given to the APPERCEPTION ENGINE are sequences of sets of ground atoms. More specifically, the atoms are of the form $senses(\alpha_i, \alpha_j)$, where α_i is a ground atom representing the *content*, constructed from a binary or unary predicate and one or two constants, and α_j is a natural number representing the time of sensation. Equivalently, Evans provides the following definition:

Definition 1. *A **sensory sequence** is a sequence of sets of ground atoms $S = (S_1, S_2, \dots)$, where S_i represents a partial description of the world at time i. Defining $\mathcal{G}$ as the set of all ground atoms we have $S \in (2^{\mathcal{G}})^*$.*

A *theory* that can explain such a sequence is defined as follows:

Definition 2. *A **theory** is a four-tuple $\theta = (\phi, I, R, C)$ where:*

- *$\phi = (T, O, P, V)$ is a type signature. T is a set of types for objects, predicates and variables, and O, P, V are sets of objects, predicates and variables which are typed according to T.*

- *I contains* initial conditions*; ground atoms representing the partial state of the world at the initial time step.*

- *R is a set of unground rules in* DATALOG$^{\ni}$*, meaning that all terms are variables. Besides normal unground* DATALOG *rules R contains causal rules: $\alpha_0 \,\overset{\frown}{\leftarrow}\, \alpha_1 \wedge, \dots, \alpha_n$, where the truth of the body at time t enforces the truth of the head at time $t+1$.*

- *C is a set of constraints grouped in three kinds. Unary constraints are of the form $\forall X\, p_1(X)\ XOR\ \dots p_x(X)$, enforcing the truth of exactly one of the p_i for each ground term that can be substituted for X. Binary constraints are of the form $\forall X \forall Y\, r_1(X, Y)\ XOR\ \dots, r_n(X, Y)$, again representing an exclusive disjunction. Uniqueness constraints are of the form $\forall X \exists! Y\, r(X, Y)$, meaning that all ground terms that can be subsituted for X are related by r to exactly one ground term that can be subsituted for Y.*

Now each theory θ generates an infinite sequence of sensory atoms called its *trace* $T(\theta) = (A_1, A_n, \dots)$. This is the minimal sequence of ground atoms such that $I \subseteq A_1$ and all rules in R are satisfied, along with one additional condition in the form of a *frame axiom*:

Definition 3. ***Frame axiom:*** *if $\alpha \in A_{i-1}$, where A_{i-1} is an element in a sequence of sets of ground atoms, and there is no atom $\beta \in A_i$ such that the truth of α and β at a single time is precluded by a constraint in C, then $\alpha \in A_i$.*

Explaining is defined as follows:

Definition 4. *A theory θ with trace $T(\theta) = (\theta_1, \dots)$ **explains** a (finite) sensory sequence $S = (S_1 \dots S_n)$ if $S \subseteq T(\theta)$, meaning that $S_i \subseteq \theta_i$ for $1 \le i \le n$.*

Now a theory is said to *make sense* if the following holds:

Definition 5. *A theory θ **makes sense** of a sensory sequence S if θ explains S, and θ satisfies the following unity conditions:*

1. *Objects are united via chains of binary relations.*

2. *Predicates feature in at least one constraint.*

3. *Every 'state' θ_i is closed under non-causal rules in R and satisfies C.*

4. Every pair of successive states θ_i, θ_{i+1} satisfies the causal rules in R.

Whether the final two conditions are in fact direct results of the definition of a trace is perhaps a matter of interpretation. I further take up the content and significance of these conditions in the following chapter. Now the *apperception task* to be executed is, given sensory sequence S, type signature ϕ and input constraints C (where ϕ provides suitable types for S and C), to find a lowest-cost theory $\theta = (\phi', I, R, C')$ explaining S, such that $C \subseteq C'$, $\phi \subseteq \phi'$. It can be shown that objects are permanent, in the sense that every object occurring in $T(\theta)$ must occur at every time step. Furthermore, for every apperception task there exists a solution. Now to solve an *apperception task*, the AE is often also given a template of the form $(\phi, N_{\rightarrow}, N_{\ni}, N_B)$, where ϕ is a type signature and the other three arguments are natural numbers representing the number of static rules, causal rules, and maximum body atoms. Templates are often provided as input for efficiency, but can also be constructed by iterating over sets of types and iterating over templates for each typeset (Evans, 2020, p.54). I now provide an example of an *apperception task* to highlight the definitions above:

Example 1. Consider the following sensory sequence:

$$\begin{aligned}
S_1 &= \{\} \\
S_2 &= \{off(a), on(b)\} \\
S_3 &= \{on(a), off(b)\} \\
S_4 &= \{on(a), on(b)\} \\
S_5 &= \{on(b)\} \\
S_6 &= \{on(a), off(b)\} \\
S_7 &= \{on(a), on(b)\} \\
S_8 &= \{off(a), on(b)\} \\
S_9 &= \{on(a)\} \\
S_{10} &= \{\}
\end{aligned}$$

To incorporate this sequence in an *apperception task*, we have to provide a suitable type signature and constraints. Now we let $C = \{\}$ and Φ as follows:

$$\Phi = \left\{ \begin{aligned}
&T = \{s\}, \\
&O = \{a : s, b : s\} \\
&P = \{on(s), off(s), p_1(s), p_2(s), p_3(s), r(s, s)\} \\
&V = \{X : s, Y : s\}
\end{aligned} \right\}$$

This suffices to define the apperception task, but we often also provide a template with upper bounds to speed up the computational process, e.g. $N_{\rightarrow} = 4, N_{\ni} = 4, N_B = 2$. An example of a theory $\theta = (\phi', I, R, C)$ that satisfies definition 5 is then a theory whith $\phi' = \phi$ and the following initial conditions, rules and constraints:

$$I = \begin{Bmatrix} p_1(b) \\ p_2(a) \\ r(a,b) \\ r(b,a) \end{Bmatrix} R = \begin{Bmatrix} p_1(X) \ni\!\!- p_2(X) \\ p_2(X) \ni\!\!- p_3(X) \\ p_3(X) \ni\!\!- p_1(X) \\ p_1(X) \rightarrow on(X) \\ p_2(X) \rightarrow on(X) \\ p_3(X) \rightarrow off(X) \end{Bmatrix} C' = \begin{Bmatrix} \forall X : s \;\; on(X) \; XOR \; off(X) \\ \forall X : s \;\; p_1(X) \; XOR \; p_2(X) \; XOR \; p_3(X) \\ \forall X : s \;\; \exists Y : s \; r(X,Y) \end{Bmatrix}$$

The trace of this theory then includes all atoms from the sensory sequence, along with missing values for the 'on' and 'off' predicates, and the atoms with predicates p_1, p_2, p_3 that define the underlying mechanism. △

Importantly, sensory sequences may also be given to the system as raw sensory data such as light intensity or pixels. In this case, a binary neural network classifies the raw data into disjunctions over ground atoms. This network is implemented within the ASP framework, providing a complementary execution of the sub-symbolic and symbolic processes. I do not make use of this full system in the current project, although it must be taken into account in the comparison of the AE with CPR made in the following chapter.

1.3 The Critique of Pure Reason and its relevance for Artificial Understanding

The Critique of Pure Reason (CPR) is a work of which the influence on modern philosophy can hardly be overestimated. While it is difficult to summarize CPR without deriving it of its subtle complexities, I here provide a brief overview of its structure and describe some of its essential concepts. A clear picture of the central aims with which Kant wrote this extensive book is helpful to guide the critical evaluation of the Apperception Engine in the next chapter.

One central aim with which Kant wrote the Critique was to for once and for all set metaphysics on the "secure path of science" (p.106); to provide it with a properly delineated subject and method, so that it would finally be able to follow the natural sciences in the reliable and progressive accumulation of insights. This required an explanation of how necessary *a priori* concepts can pertain to *a posteriori* experience, as well as a restriction of the applicability of such *a priori* concepts, preventing their unbounded dogmatic application beyond the domain of experience. Achieving both objectives simultaneously comprised a fundamental shift in perspective: to let go of the idea that our cognition must *a priori* conform to objects, and instead suppose that the objects must conform to our cognition (BXVi). If we accept that these objects, as objects of experience, are dependent on our subjective constitution, it directly becomes clear how *a priori* knowledge of them is possible. By studying the necessary preconditions for the possibility of experience in general, we can come to insights that are independent of experience yet necessarily true of anything that can appear to us. As Kant puts it:

"For experience is itself a species of cognition which involves understanding; and understanding has rules which I must presuppose as being in me prior to objects being given to me, and therefore as being *a priori*. They find expression in *a priori* concepts to which all objects of experience necessarily conform, and with which they must agree" (BXVII-BXVIII)

This is then our central interest in the Critique and its relevance to AI: the resurrection of metaphysics as a *transcendental philosophy*, meaning a science of the *a priori* conditions for the possibility of experience (A12). Importantly, this does not so much concern an investigation of our specifically human constitution, although it is necessarily bound by our own subjective reality. It is more an analysis of the necessary requirements for experiencing a world *as we do* (B72). Among other things this means a unified experience in space and time, that comprises discursive understanding as well as sensation (Longuenesse, p.12). The larger part of the Critique is an extraordinarily thorough analysis of how such experience is possible and can as such be regarded as a valuable source of inspiration for any designer hoping to build an agent that generates experience in a similar manner; that does not merely execute predefined competences or optimize parameters, but instead builds a unified world that is both sensed in space and time and represented through discursive thought.

In the remainder of this section, I provide an overview of some of the central concepts in Kant's proposal of a transcendental philosophy, laying the foundations for an evaluation of the computational implementation of this framework in *Kant's cognitive Architecture* (Evans, 2019), and paving the way for the proposal of an extended Apperception Engine; one that captures more structure from Kant's transcendental critique.

1.4 Kant's cognitive architecture

Essential to Kant's analysis of experience is the claim that it is constituted by continuous activity, as is illustrated by the following:

"Reason, in order to be taught by nature, must approach nature with its principles in one hand, according to which alone the agreement among appearances can count as laws, and, in the other hand, the experiments thought out in accordance with these principles – yet in order to be instructed by nature not like a pupil, who has recited to him whatever the teacher wants to say, but like an appointed judge who compels witnesses to answer the questions he puts to them" (BXIV)

Transcendental philosophy, as an analysis of that which necessarily pertains to the application of cognition to objects, is thus also a study of the activity that constitutes experience. An essential aspect of this activity is that it is aimed at intuition:

"In whatever way and through whatever means a cognition may relate to objects, that through which it relates immediately to them, and at which all thought as a means is directed as an end, is intuition" (A19)

Kant distinguishes several structural components in the constitution of experience by means of the faculties. A manifold of intuition is provided to us by **sensibility**, which is "the capacity to acquire representations through the way we are affected by objects" (B33-B34). Cognition or 'thought' is brought about through the **understanding**, which Kant characterizes as the "capacity to judge" (B94). The activity of combining and binding intuition pertains to the **imagination**. I now explain the role of each of these three faculties in more detail.

Sensibility

Kant states that our cognition arises from two fundamental sources in the mind:

"the first of which is the reception of representations, ... the second the faculty for cognizing an object by means of these representations" (B75)

The first is the faculty of sensibility, through which objects can be given to us. Intuition is acquired through sensibility as the way in which we are affected by objects, and hence is that which provides cognition with its content so that:

"we have no concepts of the understanding and hence no elements for the cognition of things except insofar as an intuition can be given corresponding to these concepts" (B XXVI)

The usage of phrases such as being 'affected by object' and objects being 'given to us' might mislead one to believe that Kant considers objects to be external entities that causally interact with sensibility. Such an application of causality beyond the domain of experience would however be very un-Kantian. As Béatrice Longuenesse explains, Kant considers the object of intuition as internal to its representation, not acting upon representation but making representation possible (Longuenesse, 2020, p.23). The effect of this 'object of intuition' on sensibility is sensation, and the modification of sensibility which has sensation as its matter Kant calls appearances.

Besides matter, sensibility is also characterized by its form, which is also referred to by Kant as "pure intuition". This form is that within which sensation can be ordered, so that it cannot itself be sensation. While the latter is given to us *a posteriori*, the former must in fact lie *a priori* in the mind. Without now further exploring Kant's argument in the Transcendental Aesthetic, I note that he finds two forms of sensibility: space, as the form of outer sense, and time, as the form of inner sense. These forms of sensibility are then the first example of Kant's famous reversal from object to subject: space and time, as forms of sensibility, are that which necessarily makes representation in intuition possible, and hence *a priori* structures our experience (B35). Therefore, Kant claims, we can retain and support the "apodictic certainty" of Euclidean geometry as a pure application of the spatial form of sensibility, but only so within the domain of possible experience (B41). This is Kant's distinction between empirical reality and transcendental ideality.

Understanding

The second fundamental source of cognition is understanding, which Kant describes as the faculty of bringing intuition under concepts. While through sensibility objects are given to us, through the understanding they are thought. This activity of conceptualization brings forth the representation of an object as object *corresponding to intuition*, as opposed to the indeterminate object *given in intuition* (Longuenesse, 2020, p.24). As with sensibility, Kant seeks to single out the form of understanding, as the *a priori* structure by means of which it is applied (A56). Since the understanding is the faculty of unity under concepts, its applicatory structure must be found in the functions that represent such unity; these are the Kantian judgements, hence the characterization of understanding as the 'capacity to judge'. The judgements are universal rules of discursive thought, the necessary mental activities that bring forth experience (Longuenesse, 2020, p.5). Importantly, the logical functions expressed by judgements give rise to equally many Categories. These "pure concepts of the understanding" (B105) are the unifying functions of the understanding in its application to intuition. They generally represent the necessary unity that is the Kantian object, so that Kant describes them as:

"concepts of an object in general, by means of which its intuition is regarded as determined with regard to one of the logical functions for judgement" (B129)

Synthesis, Imagination & Unity

The act of combining representations into a unity Kant calls **synthesis**. Kant applies different modes of explanation to approach this term, as is shown by the following two definitions he places in direct succession:

"the spontaneity of our thought requires that this manifold first be gone through, taken up, and combined in a certain way in order for a cognition to be made out of it. I call this action synthesis" (A77/B102).

"By synthesis in the most general sense, however, I understand the action of putting different representations together with each other and comprehending their manifold in one condition." (B103)

In the Transcendental A Deduction Kant presents the role of synthesis as an act of the imagination that combines representations, generating a sensible manifold that may be thought under concepts. In the B-deduction however, he states that synthesis is an action of the understanding and associates it with representation (B130). The first explanation emphasizes construction, while the second clarifies the relation of logical judgement to sensible syntheses. In order to understand how both explanations are complementary, it is important to analyse how Kant understands unity. In the Transcendental A Deduction he emphasizes the notion of 'unity of rule' (A105). The unification of a manifold under a concept is simultaneously a function of synthesis according to a necessary rule. For instance, our consciousness of the composition of a triangle from three straight lines according to a necessary rule of construction is what allows its representation as a unified object. For Kant, to cognize an object is to represent such a unity of rule, a unity through the function of judgement, under the concept of a category.

Now all *a priori* necessity must have a transcendental ground, hence there must also be such a ground for the unconstitutingity that represents the necessary synthesis of intuition that is expressed through the cognition of objects. This ground is the transcendental unity of apperception: the unity of consciousness that grounds all conceptual thought just as space and time ground the manifold of intuition for sensibility (A107). It is simultaneously the consciousness of one's own identity by virtue of which all experience is *my* experience and *one* experience, and the necessary unity of rule that represents a "common function of the mind" (A110) grounding all synthesis under concepts. In the same manner, the categories are the representations of unity under the functions of judgements that have *a priori* objective validity by making the very cognition of objects possible. The possibility of unity through the categories itself rests however on the transcendental unity of apperception, the unity of self-consciousness which enables synthetic unity under concepts in general (A112).

This synthetic unity in turn presupposes the activity of synthesis, which is then the effect of the imagination. Through combination and reproduction the imagination synthesizes the manifold of intuition, according to the rules represented in their conceptual unity, which is in turn grounded by the transcendental unity of apperception. As Kant puts it:

"The unity of apperception in relation to the synthesis of the imagination is the understanding" (A119)

In this thesis I refrain myself from further stretching the application of the term 'synthesis' to the operations that the AE applies to its input data. I refer to the latter as operations of combination, using synthesis in its Kantian sense as the general act of combining representations and constituting their unity.

We now have an overview of how these three faculties each play their part in constituting experience, and have seen how each comes with its own *a priori* structure. The form of sensibility, or pure intuition, grounds the totality of perception as a manifold of intuition in space and time. The pure imagination grounds the association of this manifold through synthesis. The pure apperception grounds the representation of this synthesis and its recognition as a unity in consciousness. To be bound in such necessary unity is to have objective validity, i.e. to be true of an object (A125). This then explains how necessary structure such as causality can *a priori* pertain to that objective reality that we call nature. It is the very same transcendental unity that makes experience as a unified whole possible, that legislates its necessary connection through the laws of the understanding, its objectivity.

Chapter 2

Comparing the Critique of Pure Reason with the Apperception Engine

We may now evaluate the Apperception Engine in light of what was said in the preface about understanding. The AE formulates its theories symbolically, often rendering its behavior as easily explainable as that of GOFAI. Since it constructs its own symbolic structures instead of being given them by design, it is however not as inflexible as these traditional systems. The inclusion of a neural network classifier within the ASP framework further allows the AE to handle raw and noisy sensory input, and sub-symbolic and symbolic computation are operated jointly, as is the case for Kant's faculties of *sensibility* and *understanding*. We might then say that the AE shows characteristics of Dennet's *Popperian creatures*: it is supplied with the competence to construct its own models and hypothesis using domain-general inductive biases both on the sub-symbolic and symbolic level.

This might suffice to illustrate why the Apperception Engine represents a novel approach towards AI and may serve as a starting point for an analysis of artificial understanding. In this chapter I now assess the extent to which the AE is and is not an implementation of Kantian architecture. I base my analysis on the model itself [1] as well as chapter 6 of *Kant's cognitive architecture*. The result is an overview of the essential differences between Kant's architecture and the AE, along with several proposed reinterpretations that explain the existing computational model in a manner that is closer to CPR. The intention of this analysis is not to undermine the scientific value of the AE as an implementation of CPR or a unique approach towards AI. Nor is this analysis an overview of all extensions that are feasible to implement or that I aim to implement myself. Instead, its purpose is to be used as a starting point and inspiration for anyone aiming to make the Apperception Engine more Kantian.

Let us first draw the structural similarities between Kant's architecture and the Apperception Engine in overview. The Apperception Engine receives input 'intuition', either as elementary atomic formulas relating objects to each other and to properties, or as raw sensory information such as sound frequencies or light intensity values for sensor-objects. This input is then taken up and combined into logical structures, resembling the unity under the functions of judgements in the understanding. The available logical structures are causal and non-causal rules in $\text{DATALOG}^{\ni}$, and incompatibility constraints are modelled as exclusive disjunctions. The predicates that feature in these structures are to a large extent inspired by a set of "pure operations and relations" (Evans, 2020, p.150). They are to represent

[1] https://github.com/RichardEvans/apperception

the synthesis performed by the imagination, and are identified with Kant's transcendental schemata (Evans, 2020, p.169). From the structure of these pure operations and relations Evans derives Kant's categories, thereby representing Kant's thesis that the synthesis of the imagination is structured under the unifying categorical concepts. Finally, the logical combination performed by the AE is constrained by several 'unity conditions', which are largely inspired by Kant's principles of pure understanding.

I firstly compare the structure of space and time for the AE to the forms of sensibility in CPR. Then, I compare the structure of objects in the two systems and consequently analyse all implemented unity conditions.

2.1 Sensibility, Space and Time

While the general parallels between the AE and the Kantian structure described in the previous chapter are clear, there are also significant differences between the two. I firstly consider differences regarding the nature of sensibility and the corresponding structure of pure and empirical intuition.

2.1.1 Intuition in space and time

For Kant space and time are, as pure intuitions, the *a priori* form of sensibility. Only through their agreement with space and time can appearances be given in sensibility (A92/B125). The Apperception Engine, on the other hand, does not seem to have an intuition of space and time that is *a priori*. It is given atomic formula's (or vectors from which atomic formula's are constructed) that relate objects and attributes using the pure operations and that are ordered with natural numbers representing 'subjective time'. This input is then often, but not necessarily, provided with 'objective' spatial and temporal structure using the relations of 'inheritance', 'succession' and 'simultaneity'. The Kantian manifold of intuition in space and time that is given in sensibility and taken up and combined by the imagination has thus made place for an input that can, if needed, be provided with additional spatial and temporal structure. Evans claims that this synthesis of space and time through imagination after its reception in sensibility is in line with Kant's critique:

"The job of sensibility is just to provide us with intuitions, but not to arrange them in objective space/time. It is the function of *synthesis*, the job of the imagination, to connect the intuitions together, using the pure operations and relations described above, so as to construct the objective spatio-temporal form." (Evans, 2020, p.150)

Indeed, Kant clearly states that both the existence of objects in space and time and the pure intuition of space and time themselves are the result of synthesis. In the first Analogy of Experience he writes that the existence of objects in time can only come about through their combination in *a priori* concepts (B219), hence through synthesis, and in the three syntheses of the A-deduction Kant explains how the 'pure synthesis of apprehension in intuition' is necessary for the *a priori* representation of space and time. However, the fact that for Kant synthesis is involved in the construction of space and time, does not imply that the representation of space and time as a posteriori relational constructs adequately represents Kant's architecture.

Firstly, space and time for Kant are not only *forms of intuition*: structures that are applied to empirical intuition, but are also themselves objects as *formal intuitions*. In fact, Kant claims that while space and time as *forms of intuition* provide a manifold, it is the *formal intuition* that gives "unity of the representation" (B161). These representations of space and time as objects require a pure synthesis, which is non-existent in the AE. Secondly, as forms of sensibility space and time *necessarily* structure how intuition is given in sensibility. The AE does have a succession of impressions placing all intuitions in subjective time, but this is given beforehand and not the result of an active process. The objective time on the other hand can if needed be generated through search. The structure of space can also be optionally added to intuition by the imagination after it has been given in sensibility. For the AE all structure in intuition is thus either given as input or added after reception in sensibility, so that the terms *form of intuition* and *formal intuition* do not really apply. Both distinctions are made clear by the following sections in the Transcendental Aesthetic:

"Space is a necessary representation, a priori, that is the ground of all outer intuitions. One can never represent that there is no space, though one can very well think that there are no objects to be encountered in it. It is therefore to be regarded as the condition of the possibility of appearances, not as a determination dependent on them" (B39)

"Time is a necessary representation that grounds all intuitions. In regard to appearances in general one cannot remove time, though one can very well take the appearances away from time. Time is therefore given a priori." (B46)

A close reading of Kant's explanation of the *Synthesis of Apprehension in Intuition*, more concretely shows where the two models diverge:

"Every intuition contains a manifold in itself, which however *would not be represented as such if the mind did not distinguish the time in the succession of impressions on one another*; for as contained in one moment no representation can ever be anything other than absolute unity." (A99)

Kant thus claims that synthesis in time is needed in order to represent intuition *as manifold*, as opposed to an indeterminate unity that can (paradoxically enough) not be unified to constitute experience:

"Now in order for unity of intuition to come from this manifold (as, say, in the representation of space), it is necessary to run through and take together this manifoldness, which action I call the synthesis of apprehension" (A99)

If we further note that the pure synthesis of apprehension constitutes space and time as pure intuition (A100), it is clear that while space and time are subject to synthesis they are not adequately represented as relational structures added to a manifold of intuitions given beforehand. The pure spatiotemporal synthesis that constitutes space and time also has a non-pure use in its application to empirical intuition. This synthesis then places intuition in space and time, thereby representing it *as manifold* and thus performing an essential step in the constitution of experience. I summarize the above as the following distinctions:

1. For Kant space and time as formal intuitions are the result of pure synthesis, whereas the Apperception Engine does not have a pure representation of space and time as formal intuitions.

2. The pure synthesis of time grounds the manifold of intuition for Kant, whereas the Apperception Engine is given a manifold of sensory atoms as input.

3. For Kant intuition as a manifold is necessarily represented in space and time, whereas the Apperception Engine has time and space either as input or as optional structure.

2.1.2 The structure of space and time

Above I have described differences between CPR and the AE regarding the role of space and time in the representation of intuition. I now draw attention to a related but distinct topic: the structures of space and time themselves. Besides the extent to which space and time are constitutive of the representation of a manifold as pure intuition, we can ask whether the spatial and temporal structures produced by the imagination of the Apperception Engine resemble the *a priori* intuitions that are their Kantian counterpart.

2.1.2.1 The structure of space

Space is constructed by the Apperception Engine as a set of binary 'containment' relations ('in'), forming a strict partial order with a maximal container (Evans, 2020, p.152-154). The binary relations are said to represent the transcendental schemata of the categories of quantity (Evans, 2020, 169-170). This spatial structure seems quite far from the spatial intuition Kant describes as "essentially single", an "infinite given magnitude … boundless in the progress of intuition" (A25) and containing an "infinite set of representations within itself" (B40). Firstly, space for the AE is not a single structure within which the manifold of spaces rests merely on limitation. To be fair, the unity of Kantian space is represented in the containment hierarchy by the maximal container object that is, like Kantian space, generated by the imagination in its synthesis. However, this maximal object is not itself identifiable with the spatial structure. Secondly, the spatial structure constructed by the Apperception Engine is not infinite or boundless, but specified by the objects that form its relations. That Kantian space is more complex than a containment hierarchy is also clear from the fact that it should support the three-dimensional structure of Euclidean geometry. Henkin, Suppes, and Tarski (1959) have shown that binary relations are not expressive enough to adequately represent the positions of objects in Euclidean space.

Evans clearly acknowledges the differences between the two structures, and justifies his representation by arguing that "space-qua-unifier-of-intuitions" is different from "space-qua-form-of-human-outer-sense" (Evans, 2020, p.154), and that the first boils down to the containment hierarchy. I will now investigate this claim. In the B-deduction Kant explains that the unity of space and time as pure intuitions, and hence also the agreement of intuition with space and time as forms of sensibility, stands under the general synthesis of intuition "in an original consciousness" (i.e. the transcendental unity of apperception), and thus under the categories (B161). Consequently, if a manifold of outer intuition is given in the necessary unity of space, this same unity is the category of quantity as "the category of synthesis of the homogeneous in intuition in general" (B162). In the *Schematism of pure Concepts of the Understanding* Kant then states that the image of the categories of quantity, as a "product of productive imagination" is space for outer sense and time in general, whereas its schema is number, "which is a representation that summarizes the successive addition of one (homogeneous) unit to another" (B182).

Since the containment hierarchy together with concept-determinations conveys "all the information we need for counting" (Evans, 2020, p.153), namely counting the number of objects within a container, Evans argues that this hierarchy adequately represents the schema of number. He also notes that this representation is in line with Kant's claim that space allows us to represent something as outside me, or as in different places (A23/B38). Hence, he concludes, space-qua-unifier-of-intuition is adequately modelled by the containment hierarchy

However, I argue that to separate space-qua-unifier-of-intuition from space-qua-form-of-outer-sense and to reduce the former to the schema of number is a severe simplification of the Kantian system. The sections referred to by Evans are part of Kant's *Transcendental Deduction of the universal use of the Categories in Experience*. Kant explains that the unity of apprehension in intuition (i.e. the representation of intuition as a manifold) must be in agreement with space and time as forms of sensibility. Since space and time themselves as *a priori* intuitions contain a manifold, their unity *a priori* conditions all synthesis of apprehension. As we noted before these *formal intuitions*, i.e. space and time as objects, provide unity to the manifold of intuition. Kant's aim here is then not to expound the combinatory operations made possible through space and time, but to argue that through space and time all synthesis falls under the same fundamental unity:

"Consequently all synthesis, through which even perception itself becomes possible, stands under the categories, and since experience is cognition through connected perceptions, the categories are conditions of the possibility of experience" (B161).

This however does not imply that the role of space in the synthesis of intuition can be reduced to counting. Space as form of sensibility stands under its unity as homogeneous magnitude, which in turn follows the rule of successive addition under the schema of number. The spatiotemporal synthesis under this schema thus constitutes the Kantian *a priori* intuition of space. Kant claims that synthetic unity under the category of magnitude renders appearances extensive magnitudes, meaning that "the representation of the parts makes possible the representation of the whole" (B203). For Kant, 'number' as homogeneous addition thus progressively produces space, with all the structure he ascribes to it in the transcendental aesthetic. This is not the same, and in fact the opposite, as claiming that space can be reduced to a containment structure that enables counting.

How time and space as unified *formal intuitions* play an essential role in the constitution of experience can be illustrated by an example. Consider a ball being thrown upwards. How can one produce unified experience from a succession of 'mental events', i.e., recognise a single object traversing through space and time? As a start there must be a manifold (through the *forms of intuition*), distinguishing moments in succession (A99), as opposed to a single absolute unity. The unified representation of such a manifold however, is dependent on the structure of space and time as a whole. Different moments in the balls trajectory can only be successive if they are understood as limitations of the same unique time. Similarly, we cannot cognize the succession of the ball being in place A by the same ball being in a higher place B, without these two spaces being distinguishable limitations of the same space. Furthermore, these unified representations of space and time are necessarily dependent on each other. When the ball is thrown upwards, the air it leaves below does not simply disappear, so that its space must persist through time simultaneously with the space of the ball. If we would not recognise temporal permanence in space, every spatial synthesis might be represented as a new isolated space coming into existence, rendering movement through space as a unified whole insensible. On the other hand, temporal succession as cognition of *the same thing* at different times, presupposes identity of spatial locations. If space were not represented as unity, but as a sequence of distinct spaces, there would be no ground for the experience that the ball being in place A is succeeded by the ball not being in the same place A at the next moment.

We thus see that unifying intuition requires the representation of space and time as *formal intuitions*. No similar synthesis can be grounded by mere counting or a containment hierarchy. One might, using a containment hierarchy, represent a ball as being in one space A, and on another occasion represent a ball as being in a space B, but if these containers both contain the same objects (a ball and air) they might in fact be two representations of the same space. Even if we assume the containers are distinct, there is no representation of their relative position as limitations of space as a whole, so that there is no structure to represent continuous motion between them.

I have argued that space-qua-form-of-outer sense is not separable from space-qua-unifier-of-intuition, since the same unifying synthesis that generates this form is presupposed in the synthesis of empirical intuition. In fact, following Béatrice Longuenesse, we might consider the synthesis of space to be even more deeply connected to the empirical synthesis of experience. She argues that all three syntheses of the A-deduction should be understood as belonging to the same act with a threefold outcome (Longuenesse, 2020, p.35). The synthesis of apprehension in intuition is "inseparably combined" (A102) with the synthesis of reproduction in imagination. Now the synthesis of recognition in the concept expresses the unity of act that grounds the consciousness of the generic identity of representation. Kant states that without this consciousness "all reproduction in the series of representation would be in vain" (A103). We can thus understand the three acts of synthesis as inseparable. Following Longuenesse we can interpret this description of a threefold act as exhibiting an apprehension, a reproduction, and a recognition for each *particular intuition*. The form of this threefold synthesis is then provided by the same pure spatiotemporal synthesis (Longuenesse, 2020, p.47). An interesting representation of this idea is provided by the Figurative Apperception Engine introduced in chapter 4.

Whether or not one follows this interpretation all the way, I believe it is safe to conclude that Kant considers the formal intuition of space to be essential for the synthesis of intuition. Understood in this way, everything Kant states about space in the Transcendental Aesthetic directly applies to space-qua-unifier-of-intuition, including his claim that "Space is not a ... general concept of relations of things in general" (A25), as is the case for the containment hierarchy. I thus add the following two distinctions between the two models:

5 For Kant the same synthesis underlies space-qua-form-of-outer-sense and space-qua-unifier-of-intuition, whereas the Apperception Engine only represents the latter.

6 For Kant space is essentially single, boundless, infinitely divisible and supportive of Euclidean Geometry, while space for the Apperception Engine lacks all these traits.

2.1.2.2 The structure of time

The Apperception Engine is given a sequence of 'subjective times' as natural numbers associated with each sensory atom. This subjective time can be converted into natural numbers of 'objective time' based on the relations of 'simultaneity' and 'succession', where successive subjective times can be placed either in simultaneous or successive objective times, and the atoms in objective time must be consistent with the constraints of causality and incompatibility the Apperception Engine produces. As was the case for space, the structure of natural numbers is different from the "necessary representation that grounds all intuitions" (A31) Kant describes in the Transcendental Aesthetic. He emphasizes the essential singularity of time, so that "every determinate magnitude of time is only possible through limitations of a single time grounding it"(B48). Such an *a priori* structure representing time as a whole is not found in the Apperception Engine, and its magnitudes of time are hence not formed by limitation but through generation of individual times. This is clearly not in line with how Kant considered the synthesis of time, since he claims that "from mere places, that could be given prior to space or time, neither space nor time can be composed" (A170). Kant also describes time, like space, as an infinitely divisible quantum continuum (B211), which the natural numbers are not.

As is the case for space, the binary relations that generate time for the Apperception Engine, namely 'simultaneity' and 'succession', are based on transcendental schemata, in this case these are the schemas of cause and community (Evans, 2020, p.170). What Evans does not note is that for Kant the generation of time, like that of space, falls under the schema of quantity. While space is the image of all magnitudes for outer sense, time is the image "for all objects of the senses in general" (B182). This correspondence between time and quantity is in fact to some extent represented in the AE since it constructs time as countable numbers (to the same extent that quantity in space is represented by a containment structure that enables counting). The matter becomes more complicated since Kant does state that time as *form of intuition* contains relations of succession and simultaneity (B68). It thus seems that while the categories of *quantity* for Kant ground the synthesis of time itself, the application of time as form in empirical synthesis is characterized by the relations chosen by Evans. The apparent distinction between the relevant categories for temporal synthesis is then a consequence from what was expounded in section 2.1.1: time for the AE is only generated *a posteriori* from empirical intuition. As we have concluded before, the synthesis of apprehension in intuition that generates time in its pure application and that distinguishes "the succession of impressions" (A99) is not present in the structure of the AE. Instead, the structure of space and time is either given beforehand or generated as relational construct *a posteriori*. Given this distinction it is then sensible that the relation of 'containment' is not used in the AE for the generation of time, but that instead the generation of time itself as homogeneous magnitude is modelled by a numerical structure in the input. It must be noted again however that this is quite un-Kantian. For Kant the succession in impressions is not given as an input-sequence, but instead generated by the act of apprehension itself (Longuenesse, 2020, p.37).

I now briefly explain why, given that the AE does not have a synthesis of apprehension, its relational representation of time through simultaneity and succession is in line with CPR. In the same *Transcendental Deduction of the universal use of the Categories* where Kant states that the synthesis of apprehension in outer intuition presupposes the spatial synthesis under the categories of quantity, he also refers to an example of temporal succession, namely the freezing of water. Kant claims this succession can only be determinately given if it is grounded on the synthetic unity of time as inner intuition. However, Kant does not refer to quantity here as the generation of time itself, but instead refers to the category of cause through which "I determine everything that happens in time in general as far as its relation is concerned" (B163). This example is in line with Kant's introduction of the Transcendental schemata, where he states that the schemata of relation, including the schema of cause and that of community, contain "the relation of the perceptions among themselves to all time" (B185). Since the Apperception Engine generates the structure of time from the temporal relations between objects, and since these temporal relations for Kant fall under the schemata of relation, the use of simultaneity and succession is quite reasonable. As was said before, this *a posteriori* construction of time is however different from the Kantian system where simultaneity and succession can only come into perception under the presupposition of time as *a priori* intuition (A31). Furthermore, the activity of relating objects using causality and simultaneity presupposes representations of objects. For Kant the construction of such representations must itself be grounded in synthesis under the categories and, again, the unity of *formal intuitions*. This is further explained in the following section. I summarize the above in two distinctions:

7 For Kant the generation of time through a pure synthesis of apprehension is grounded in the schema of number, while for the Apperception Engine succession in time is given as input and all construction of objective time follows the schemata of cause and community.

8 For Kant time is boundless, infinitely divisible and represented as a whole, while time for the Apperception Engine lacks all these traits.

2.2 Unity and Objects

An important representation of Kantian structure in the Apperception Engine is the implementation of various 'unity conditions', which are largely inspired by the *Synthetic Principles of Pure Understanding* that Kant introduces in the *Doctrine of the Power of Judgement*. In this section Kant aims to systematically describe those *a priori* judgements that the understanding brings about in the relation of the categories to sensibility, and that ground all other cognitions (B188).

As we have noted before, all synthetic unity of intuition in experience rests on time as form of inner sense, the imagination and the transcendental unity of apperception. These are three conditions for the representation of objects as objects of experience. Now for a cognition to have objective validity, Kant argues, is for it to be necessary for all objects of experience, and hence for experience in general. "The possibility of experience is therefore that which gives all of our cognitions *a priori* objective reality" (A156). The *a priori* judgements grounding all cognition are to thus be found in the conditions that ground the unified synthesis of intuition, so that the supreme principle of synthetic judgement is "Every object stands under the necessary conditions of the synthetic unity of the manifold of intuition in a possible experience" (A158). It is thus fitting that Evans has provided the Apperception Engine with 'unity conditions'. The principles of understanding are precisely this: *a priori* principles conditioning the possibility of experience and hence the synthetic unity of intuition. There is however a large difference in interpretation, since for Kant these principles only have objective validity by virtue of their being necessary for objects of experience. For the Apperception Engine on the other hand, objects are atomic entities, so that the unity conditions ground the possibility of unified experience, but not that of unity in the experienced objects themselves. Before I evaluate the various unity conditions, I must thus explain the Kantian object in more detail.

2.2.1 Objects

In the introduction of the faculties we have already identified the cognition of objects with the necessary synthesis of intuition, claiming that 'objective reality' is legislated through the laws of the understanding and the unity of apperception. I distinguished, following Longuenesse's interpretation, two aspects of the object as internalised within its representation: the given object as appearance making its representation possible and the object as object of experience made possible by its unified representation.

It is clear then that the Kantian object is very different from that of predicate logic, in which a domain of atomic objects enables variable substitution. The Kantian object is both given and thought, both internal in its representation and yet conditioning truth (B197). Why does Kant hold such a complex conception of objects, and how can we understand this conception within the computational framework of the Apperception Engine?

From our exposition of Kant's central aims for the Critique in Chapter 1, it is clear that he cannot take objects to be the atomic domain elements of predicate logic. Kant sought to provide an explanation of how *a priori* concepts can relate to the objects they represent. Any dependency of cognition on external objects would however render it *a posteriori* and hence rob it of its necessary character. Kant thus took the object to be internal to representation, and consequently needed to explain how such an internally represented object can exist and why it seems to present itself as distinct from its representation:

"What does one mean, then, if one speaks of an object corresponding to and therefore also distinct from the cognition? It is easy to see that this object must be thought of only as something in general = X, since outside of our cognition we have nothing that we could set over against this cognition as corresponding to it" (A104)

The answer to the question that Kant provides is that this general X that seems to be related to our cognition is precisely what determines this cognition as necessary. For a cognition to relate to an object is for it to "have that unity that constitutes the concept of an object" (A105). Since there cannot be an object outside cognition causing this unity, it must be grounded in the formal unity of consciousness, i.e., the unity of apperception. We hence again come to the notion of 'unity of rule', since the objectivity of cognition is its unity as synthesized under the rules of the understanding. It is from this viewpoint that Kant can define an object as "that in the concept of which the manifold of a given intuition is united (B137)".

How then does Kant explain that this internally represented necessity presents itself as an object outside or behind its representation? Why does Kant not only refer to the object *as object of experience*, but also to the object that is *given in intuition*? One explanation for this is given in the A-deduction. Appearances are themselves representations. Each representation has an object, which is often another representation. Appearances however do not represent other representations, but are "given to us immediately", so that their object "cannot be further intuited by us" (A109). This object represented as 'outside us' Kant calls the non-empirical transcendental object. This transcendental object is what ultimately provides our empirical concepts their relation to an object. It does not itself contain any intuition, but is instead again the formal unity that characterizes all objectivity, i.e., the necessary unity of apperception. It is the *a priori* function of synthesis that conditions the object of experience. One might object that the object *as appearance* in intuition must stand under the forms of sensibility, since Kant identifies these forms as the condition under which objects can be intuited (A93). I believe however that the apparent distinction between the two conditions is resolved by what was said in section 2.1.2 regarding the three syntheses of the A-deduction. They can be considered modi of the same process, providing a threefold outcome and all presupposing the general identity of act that is the transcendental unity of apperception. Hence, while appearances are given in space and time as forms of sensibility, the synthesis of apprehension providing the object as appearance in space and time is grounded in the unity of consciousness; the transcendental object.

Summing up, we see that the object of experience must be internal within its representation as a necessity or 'stable coherence' grounding its unity under concept. The object of appearance is given immediately in sensibility, but this is in turn only possible by virtue of the transcendental object as unity of consciousness under the rules of which all appearances in their relation to objects must stand. The distinction between this model and the nature of objects in the Apperception Engine is then rather clear. The Apperception Engine receives some *objects as appearances* such as sensors and numerical values in its input, and assumes the existence of other objects in order to construct a unified model of experience. The objects in the latter category are in a sense dependent on the unity of consciousness. However, the assumption of entities to facilitate model construction is not the same as the representation of intuition *as objective* by virtue of a unifying function of the mind. The unifying conditions in the Apperception Engine can bring forth objects, but are not themselves identified with a transcendental object. The Kantian *object of experience* also does not seem to have a direct representation in the Apperception Engine. On the one hand, the AE subsumes objects under types, associates types with predicates, and connects predicates in xor constraints. In this sense, the objects of the AE are embedded in a necessary structure. On the other hand, these objects are themselves represented by atomic constants that can be given in the input directly. The AE must learn which xor constraints are associated with predicates, and if the template is not given it must even learn which types are associated with which objects in the input. However, the objects themselves are still mainly a starting point for synthesis, instead of the achievements of synthesis under necessary connection of judgement. This gives the following distinctions:

9 For Kant appearances can be considered *objects as appearances* under the forms of space and time and the transcendental unity of apperception, while the Apperception Engine either hypothesises the existence of objects or is given objects as input.

10 For Kant the *object of experience* represents synthesis of intuition as necessary unity, while the Apperception Engine starts synthesis from objects as atomic entities.

2.2.2 Unity conditions

From the above it is clear how the Unity Conditions, conditioning experience as the connection between atomic objects by means of relations, differ from the principles of understanding, conditioning experience in general and equivalently the constitution of experienced objects. Having made this distinction I now evaluate in turn each of the Unity Conditions in relation to its associated principle. The differences regarding the nature of objects are a guiding thread through the differences between the two systems, but can now often be treated succinctly.

Kant divides his principles of understanding into those that are *mathematical* and those that are *dynamical* (B200). The first group directly follows from the application of the categories as conditions of intuition, while the second group follows from their application to appearance in general. The first type of principles is thus 'unconditionally necessary' while the second is only *a priori* necessary under the condition of empirical thinking in general, i.e., the synthesis of intuition under empirical concepts. Evans applies this distinction to the pure relations that represent the schemata and notes, following Kant's distinction between composition and connection (B201), that "The mathematical relations control the arbitrary synthesis of homogeneous elements, while the dynamical relations control the necessary synthesis of heterogeneous elements" (Evans, 2020, p.152). The guiding principle for the unity conditions is thus that each principle of pure understanding, as condition of experience springing from the application of a category, can be represented as a condition on the relation that represents the schema of the associated category. The mathematical relations are then 'containment' and 'comparison', as associated with the categories of quantity and quality. The dynamical relations are 'inherence', 'succession', 'simultaneity' and 'incompatibility', as associated with the categories of relation and modality.

Mathematical conditions

The first and fundamental unity condition Evans introduces for all mathematical relations is that the intuitions must form a fully connected graph. This seems to be based on the earlier mentioned 'supreme principle' Kant phrases as "Every object stands under the necessary conditions of the synthetic unity of the manifold of intuition in a possible experience" (A158). Although which Kantian principle is used as inspiration for this condition is not wholly clear to me.

I have explained above how the statement that objects stand under conditions of synthetic unity is different from the statement that objects are united in a connected graph. Kant does not identify unity of consciousness with a network of connection. Instead, it is described as "identity of action" and "the necessary unity of the synthesis" (A108). This identity of action can be associated with the function of judgement, grounding universally represented combination under the categories and even the structure of general logic. From this perspective, the very fact that all data is structured in DATALOG$^{\supset-}$ and can be combined in the general logical structures of rules and incompatibilities represents the principle of transcendental unity, by allowing the unified representation of the data under universally applicable functions.

An additional issue with this condition of connectivity is that its implementation in the AE does not match with the mathematical relations, as is stated in *Kant's Cognitive Architecture*. While the condition is implemented directly for the 'containment' relation, it is also implemented for the 'inherence' of objects under concepts. Although 'inherence' is a dynamical relation, connectivity thus enforces that all objects are connected by relational predicates. Conversely, the 'comparison' relation is mathematical, but is encoded directly as a set of mathematical axioms applied to number objects. Although this mathematical relation forms a connected graph, it thus strictly speaking not constrained by a unity condition.

The second mathematical unity condition is that the spatial structure has a maximal container. The extent to which this represents the unity of Kantian space has been expounded in section 2.1.2. I note here however that Kant's principle associated with the application of magnitude in the *Axioms of Intuition* reads "All intuitions are extensive magnitudes". This implies that intuition in space and time is extensive in the sense that its parts make possible and precede the representation of the whole. A representation of this principle is not implemented in the Apperception Engine.

The third mathematical unity condition is that the pure relation of comparison ('<') forms a strict partial order. Evans associates this relation with the categories of quality, so that "X falls under the pure concept of reality if there exists an intuition Y such that $Y < X$" (Evans, 2020, p.170). Kant on the other hand continues his exposition of mathematical principles in the *Anticipations of Perception* with the following principle resulting from the categories of quality: "In all appearance the real, which is an object of the sensation, has intensive magnitude, i.e., a degree". He argues that the 'real', can be raised from its negation (= 0) to any degree, so that between reality in appearance and its absence there is always an intermediate sensation. The associated schema for Kant is the "continuous and uniform generation of ... quantity in time" (B183). It thus not in line with the Kantian system that the unity condition for the relation of comparison merely restricts this relation to be a strict partial order, and does not insist that this comparison relation is dense (meaning that $Rxy \rightarrow \exists z Rxz \wedge Rzy$). This difference is again a consequence from the distinction between the two systems regarding the nature of objects: since the intuition that is 'compared' with the < relation consits of objects and relations instead of appearances with sensation as matter, a dense comparison relation would result in an infinite domain and intractable computation. Evans argues that this structure is in line with CPR, since Kant intended finite models. It is important to note however, that for Kant a continuum was not considered to be built up from individual points. He was thus quite consistent in both insisting that one cannot experience infinitely many objects, and that one can always find an intermediate quality between two qualities *ad infinitum* (Pinosio, 2017, p.21). I add the following distinctions:

11 For Kant unity of consciousness is represented by identity of act, whereas the Apperception Engine has an identity of act through its computational structure, but represents unity as a connectivity constraint between objects.

12 For Kant the application of quantity to intuition renders appearances extensive magnitude, whereas the Apperception Engine does not represent extensive magnitudes.

13 For Kant the application of quality to intuition provides sensation with an intensive magnitude, whereas the Apperception Engine represents intensity by a non-dense strict partial ordering of objects.

Dynamical conditions

The guiding principle Evans uses for the dynamic unity conditions is that synthesis by the imagination must be backed up by conceptual judgements. Positions of intuitions in determinations (i.e., 'containment', 'comparison', and 'inherence') "can only be fixed by *forming a judgement that necessitates this particular positioning*". Similarly, the relative positions in relations between determinations (i.e., 'simultaneity', 'succession' and 'incompatibility') can only be fixed by complex judgements that themselves relate two judgements (Evans, 2020, p.159). This principle then translates into several conditions that explain how the atomic formulas in the trace of a theory must be 'backed up' by the logical structures in the theory itself. The purpose of this structure is not fully clear to me. both from a computational and a philosophical perspective. From a computational perspective it seems that since datalog$^{\ni}$ theories generate a trace, the elements of this trace are by construction 'backed up' by the corresponding logical structures in the theory, so that the formulation of multiple dynamic constraints is an unnecessarily complicating addition. One could equivalently merely define the logical structures themselves. From a philosophical perspective it seems that the principle that is used to justify the dynamical unity conditions is formulated in a manner that is not wholly consistent with CPR. I now explain the latter and subsequently propose a reformulation of this principle that merely points to the existence of logical functions in the understanding.

On the one hand, the statement that the structure of judgement represents necessity in synthesis is very much in line with the *Critique*. As I have discussed earlier, the Kantian object of experience is identified with unity in cognition that is grounded in the unity of apperception, and as Kant states in the B-deduction: "a judgment is nothing other than the way to bring given cognitions to the objective unity of apperception" (B142). However, the structure of judgement does not, as Evans claims, necessitate the specific positions of intuitions and determinations within their relational structure. Kant emphasizes that judgements cannot make empirical cognitions necessary, since these cognitions remain *a posteriori* (B142). To use a Kantian example: it is not the case that a causal judgement teaches us that a pillow is necessarily dented by a leaden ball instead of the other way around. However, the structure of judgement does provide us with an *objective relation*, so that we can say that the 'the pillow is dented by the ball' instead of merely associating perceptions. Again, the constitution of objects as opposed to associations is not represented in the APPERCEPTION ENGINE, so that the interaction between synthesis and the conceptual unity of the understanding had to be reformulated. However, there is another aspect where this principle of dynamic unity seems to deviate from the structure of the *Critique*, prompting its reformulation.

Evans presents the imagination as constructing a synthesis, which is then provided necessity by being 'backed up' by an independent understanding:

"All the imagination can do is connect the intuitions using the pure relations – it cannot impose necessity on those connections. In fact, the only element that can provide the desired necessity is the judgement." (Evans, 2020, p.156)

From this viewpoint, it seems sensible that the dependence of synthesis on judgement is only formulated as a principle for dynamic relations. Kant distinguishes the dynamical categories as having 'correlates' (B110), which is not the case for mathematical categories. However, it is also rather clear that in fact Kant thought that *all* synthesis under unity of the understanding obtains its objectivity through the structure of judgement. This directly follows form Kant's exposition of the schematisms of pure concepts of the understanding, which are pure syntheses in accordance with (all) the categories. This then shows that the interaction between judgements and synthesis cannot be fully captured as relations being backed up by judgements. In the previous chapter I emphasized that the imagination can only perform its synthesis guided by the rules of unity. In fact, we saw that Kant stated that "The unity of apperception in relation to the synthesis of the imagination is the understanding" (A119). All synthesis, as combination aiming towards objective unity is thus also an act of the faculty of understanding, so that this faculty is not merely one of discursive relations between concepts: "all combination, ... whether it is a combination of the manifold of intuition or of several concepts, ... is an action of the understanding, which we would designate with the general title synthesis" (B130). Taking this aspect of understanding seriously would imply that the synthesizing relations used by the Apperception Engine are themselves brought forth by the understanding, whether they have 'correlates' or not. I thus propose the following principle as a substitution for the dependency of dynamic relations on judgement:

P1: All unity of synthesis of intuition is represented under the logical functions of the understanding.

This principle has a bidirectional interpretation. The first is that the synthesis of intuition is generated as the trace of a unified conceptual theory in the understanding. The second is that all synthesis by the imagination structures the sensory atoms so as to allow their unified representation in the theory. Given this principle all that is left is to specify the logical functions. This removes the need for further dynamic unity conditions, and simultaneously applies to relations of mathematical as well as dynamical nature.

Interestingly enough, this interpretation is consistent with the computational implementation of the AE. The pure operations and relations used in the synthesis of intuition are associated with the transcendental schemata. The conceptual judgements brought forth by the understanding do not fix the positions within these relations, but rather generate them as trace in their application to intuition. We can then with Kant say that the understanding "is not merely a faculty for making rules through comparison of appearances; it is itself the legislation for nature" (A127). In the remainder of this section I explain to what extent the relations featuring in the dynamic unity conditions adhere to this principle. I do not analyse the logical functions themselves, since this is the topic for the next chapter. I must also note that the representation of logical atoms under the conceptual structures of a theory does not represent Kant's argumentation in the *Analogies of experience*, where he explains how alteration of substance through causal interaction constitutes objectivity and the filling of time. This has partially been covered before and will not be discussed for each condition of dynamical unity.

Inherence

Evans formulates the first condition of dynamic unity as "inherence must be backed up by categorical judgement" (Evans, 2020, p.158). Applying **P1** instead, it is clear that every determination of an object or pair of objects by a conceptual predicate is indeed represented under a conceptual structure in the theory, containing a variable that has this object in its domain (note that all variables in rule-heads are universally quantified in $\text{DATALOG}^{\ni}$) . The relation of 'inherence' thus adheres to **P1**.

Succession

The second condition of dynamic unity is that "Succession must be backed up by a causal judgement" (Evans, 2020, p.159). I firstly argue that this unity condition is strictly speaking not satisfied by the implementation of the AE. Secondly, I claim that this implementation is not at odds with CPR, so that the relation of 'succession' can be explained as adhering to **P1**.

To see why it is not necessarily the case that all successions must be backed up by causal judgements, we can read the following ASP rule that is used by the AE to generate succession and simultaneity for any two sensed 'intuitions' in successive 'subjective times' (ST):

```
1 { sim((BV1, Obj1, ST), (BV2, Obj2, ST+1)) ;
    succ((BV1, Obj1, ST), (BV2, Obj2, ST+1)) } 1 :-
        bv_at(BV1, Obj1, ST), bv_at(BV2, Obj2, ST+1)
```

For any two sensory inputs in successive 'subjective times' a nondeterministic choice is made between simultaneity and succession in objective time. Other clauses then ensure that these determinations are assigned to numbers of 'objective time', and subsumed under the ground atoms that are the trace of the theory. Some of the successions (either on the level of determinations, or on the level of ground atoms) are directly fixed by the extension of causal rules. However, other successions might just be generated in accordance with theory optimization and the conceptual structure as a whole. This is due to the fact that the condition of subsumption by the trace of the theory is not applied to the succession relations themselves, but only to the resulting atoms in 'objective time'. For instance, if the sequence in 'subjective time' $P(a), R(a,b), P(b), \neg P(a), \neg P(b)$ is given, then $P(x)$ and $\neg P(x)$ for $x \in \{a, b\}$ cannot both be true at the same objective time. One might place the first three atoms at objective time 1, and the latter two atoms at time 2, providing the initial conditions $R(a,b), P(a), P(b)$ and the following causal rule $P(x) \ni\!\!\!- \neg P(x)$. Note however that this judgement does not directly 'back up' the succession of $P(a)$ and $R(a,b)$ by $\neg P(B)$, or the succession of $R(a,b)$ by itself. These successions are instead generated in order to *prevent* the need for more complex conceptual structures.

I now argue that this synthesis of succession by the imagination and its limited representation under causal rules is in fact in line with Kant's writing on consciousness. A first interesting fact on this topic is that Kant distinguishes representation in general (*representatio*) from representation with consciousness (*perceptio*) in the introduction to the *Transcendental Dialectic* (A320). This is perhaps not very surprising, given that Kant uses the term 'representation' widely, but insists that a synthesis of apprehension is necessary to distinguish representations in intuition from 'absolute unity' (A99). What might be more surprising is that Kant has also distinguished representations with consciousness in general from representations that are thought. In the Jäsche Logic Kant claims that animals *perceive* representations and identities and differences between them, but are not conscious of their own acts, so that they may know, but not cognise, i.e., form concepts (Kant, 2001, Logik introd. VIII, Ak. ID, 65-570). Longuenesse argues that this distinction can plausibly be applied to our own representations, so that while all intuition can *possibly* be reflected under the discursive unity of thought, this complete synthesis need not always be *actualized* (Longuenesse, 2020, p.66).

Applying this distinction to the mechanism of the Apperception Engine, we see that it is quite Kantian that combinations are made by the imagination that are not directly supported by conceptual structure. All succession can *possibly* be directly determined by causal rules, but this is often not necessary. We can hence with a bit of free interpretation associate the causal rules with those successions that the Apperception Engine thinks, while those successions that are merely generated by the imagination but not represented explicitly can be considered those of which the AE is only conscious. We can then still consider all unity of the synthesis of intuition in 'objective time' as being represented under the logical structure of the understanding, since the atomic formulas in successive objective times are directly subsumed by the trace of the conceptual theory and since the synthesis of succession by the imagination structures perception in a manner that presupposes its conformity with universal causal rules. We can thus claim that 'succession' still adheres to **P1**.

Simultaneity

The third condition of dynamic unity is "Simultaneity must be backed up by a pair of causal judgements" (Evans, 2020, p.159). Since Evans explicitly notes that this condition is not implemented in the AE, I do not dwell on it. I merely note that the simultaneity of sensory atoms can be argued to be unified under representations in the understanding by means of the same argument that was applied to succession, so that 'simultaneity' can be said to adhere to **P1**. The correctness of the association of simultaneity with a 'pair of causal judgements' is a topic that is not taken up here.

Incompatibility

The last condition of dynamic unity is "Incompatibility must be backed up by disjunctive judgement" (Evans, 2020, p.160). As was the case for inherence, every incompatibility between sensory atoms is directly represented by a conceptual incompatibility in the theory. This relation thus also adheres to **P1**. Whether or not disjunctive judgements are adequately represented as incompatibilities is again a topic that is not taken up here.

2.2.2.1 Making concepts sensible and Conceptual unity

The last two unity conditions are neither mathematical nor dynamical. They read: "judgements are supported by corresponding determinations" and "every concept features in some disjunctive judgement" (Evans, 2020, p.161-162). The first condition is not directly related to the principles of understanding. It enforces that if the AE conceptually derives an atomic formula applying a predicate P to an object o, it must also construct a corresponding 'inherence' relation on the level of synthesis by the imagination determining that an attribute a holds of o, where a falls under P. This condition is supported by Kant's claim that it is necessary to "make the minds concepts sensible" and "add an object to them in intuition". We have seen that for Kant an object is acquired through necessary unity in the synthesis of a manifold. These statements then mean that a concept as discursive unified representation does not yield cognition if it does not represent a corresponding synthesis of a manifold in intuition, which is quite different from the invention of an atomic attribute as extension of a conceptual predicate. Still, the general idea that intuition is *reproduced* by the imagination to ground the discursive unity of the understanding is in line with the *Critique*.

The second condition enforces that all concepts that are not provided in the input, and thus invented by the AE itself, must be connected to other concepts by exclusive disjunctions. This condition is inspired by the third analogy of experience, in which Kant explains that community conditions simultaneous existence of objects, and also their existence as objects of experience in general (B258). This principle of community however, applies to all appearances, as standing together in experience, and comprises a reciprocal grounding relation as well as exclusion. The unity condition on the other hand only applies to concepts, instead of appearances in relation to their objective synthesis, and only specifies some subgroups of these concepts that form an exclusive disjunction. It is then clear that the unity condition only partially captures Kant's intended structure, but a more detailed analysis of judgement must be made in the next chapter before this distinction can present itself in full clarity. This analysis gives us one final distinction:

14 For Kant all appearances in experience stand in a community where the objects stand in a reciprocal grounding relation, while the Appereption Engine represents this community by exclusive disjunctions between non-input concepts.

2.3 Conclusion

This then concludes my evaluation of the Apperception Engine in comparison with Kant's framework with respect to the forms of sensibility, the nature of objects, the interaction between imagination and understanding and the unity conditions. One major topic has been intentionally left out: the structure of judgements as unifying functions of the categories. In the next chapter I further analyse the logical structure of Kantian judgement, showing where it digresses from that of the rules and constraints in $\text{DATALOG}^{\ni}$, and proposing extensions of the AE that bridge this gap. I further note that the provided comparison is by no means a complete overview. Those who are interested in further development of the Apperception Engine might consider taking up one of the many remaining topics such as the power of judgement, the structure of modalities or the nature of reason as actively guiding the understanding according to principles. For now I content myself with the following list of distinctions between the two models, and proceed with a logical analysis of Kantian judgement to substantiate my own implementation of Kant's cognitive architecture.

1. For Kant space and time as pure intuitions are the result of pure synthesis, whereas the Apperception Engine does not have a pure intuition of space and time.

2. The pure synthesis of time grounds the manifold of intuition for Kant, whereas the Apperception Engine is given a manifold of sensory atoms.

3. For Kant intuition as a manifold is necessarily represented in space and time, whereas the Apperception Engine has time and space either as input or as optional structure.

4. For Kant the same synthesis underlies space-qua-form-of-outer-sense and space-qua-unifier-of-intuition, whereas the Apperception Engine only represents the latter.
5. For Kant space is essentially single, boundless, infinitely divisible and supportive of Euclidean Geometry, while space for the Apperception Engine lacks all these traits.
6. For Kant the generation of time through a pure synthesis of apprehension is grounded in the schema of number, while for the Apperception Engine succession in time is given as input and all construction of objective time follows the schemata of cause and community.
7. For Kant time is boundless, infinitely divisible and represented as a whole, while time for the Apperception Engine lacks all these traits.
8. For Kant appearances can be considered *objects as appearances* under the forms of space and time and the transcendental unity of apperception, while the Apperception Engine either hypothesises the existence of objects or is given objects as input.
9. For Kant the object of experience represents synthesis of intuition as necessary unity, while the Apperception Engine has objects as unstructured atomic entities.
10. For Kant unity of consciousness is represented by identity of act, whereas the Apperception Engine has an identity of act through its computational structure, but represents unity of consciousness as a connectivity constraint.
11. For Kant the application of quantity to intuition renders appearances extensive magnitude, whereas the Apperception Engine does not represent extensive magnitudes.
12. For Kant the application of quality to intuition provides sensation with an intensive magnitude, whereas the Apperception Engine represents intensity by a non-dense strict partial ordering of objects.
13. For Kant all appearances in experience stand in a community where the objects stand in a reciprocal grounding relation, while the Appereption Engine represents this community by exclusive disjunctions between non-input concepts.

Chapter 3

Judgement and Geometric Logic

3.1 Introduction: judgements as rules

In the previous chapter I have hinted at the structural differences between rules in DATALOG$^{\ni}$ and Kantian judgements. The contents of this chapter are to a large extent separate from that of the later chapters, so that the reader may choose to start with the analysis of time in chapter 4 without losing his or her way. We have seen that Kant understands judgements as functions of the mind; rules that bring representations under necessary unity. Through the structure of judgement propositions obtain their 'objectivity': their relation to truth. This mostly considered *transcendental logic* as the science determining the origin and objective validity of cognition. I now turn more towards the subject of *general logic*: the science of the logical form of reasoning that abstracts away from content and truth. An alternative definition of judgement, provided in the *Prolegomena*, is then of relevance:

"Judgements, when considered merely as the condition of unification of representations in a consciousness, are rules." (Kant, 1966, 23)

Kant has defined rules as "assertions under a universal condition" (Logik, §58, Ak. IX, 121; 615), so that judgements can in turn be understood as assertions under a universal condition. A second definition of importance is then that of *inference*:

"An inference of reason is the cognition of the necessity of a proposition through the subsumption of its condition under a given universal rule."

As an example, we can consider an instance of modus ponens where the proposition *'metal is divisible'* is inferred from the more general rule *'bodies are divisible'*. In this case, *'metal'* serves as a condition, subsumable under *'body'* so that *'metal is divisible'* obtains a necessary character through the universal rule *'bodies are divisible'*. *'Metal'* is a condition by virtue of this very possibility of subsumption. However, also the term under which is subsumed (*'body'*), must be considered as containing a condition since it enables the subsumption of 'metal', and can in turn itself be subsumed (A322). We see then that *assertion under condition* and inference are not distinct from one another. The possibility of inference allows for the cognition of judgement, and judgements can in a way be understood as *recipes for inference* (Achourioti, van Lambalgen, et al., 2017, p.856). Importantly, the concepts subordinated to each other in this manner must also have an object (x):

When I say 'A body is divisible', this means: 'some x, which I cognize through the predicates that together constitute the concept of body, I also think through the predicate of divisibility'. (Refl. 4634 (1772-76), Ak. XVII, 616.)

Thus, Kant also names judgement "the mediate cognition of an object" (A68), relating concepts to other representations (which are either other concepts or intuition) in a way that allows the representation of some x. The logical form of judgement must thus facilitate this cognition of objects under concepts trough subsumption of representations. In the next section, I follow an argument given by Achourioti and Van Lambalgen (2017; 2011), who argue that *geometric logic* adheres to this conception of judgement. Subsequently, I describe an implementation of *geometric logic* using non-deterministic choice in ASP.

3.2 Geometric Logic

The term *Geometric logic* arose in algebraic geometry, where it includes higher-order logical structure (Blass, 1993). In a first-order context the term usually refers to a logic of universally quantified implications, where the antecedent and consequent only contain the connectives $\exists, \vee, \wedge$, along with $\bot$. Infinite disjunctions are often allowed, in which case the term *coherent logic* is used for the restriction to finite disjunctions (Vickers, 1993). In a computational context, *geometric logic* or GEOLOG also applies to a structure of logical programming, which naturally does not allow for infinite disjunctions (Fisher & Bezem, 2007). This is the interpretation I hold here, which gives rise to the following definitions[1]:

Definition 6. *Given a first-order language $\mathcal{L}$ with equality, a formula is geometric in $\mathcal{L}$ if its is constructed from atomic formulas in $\mathcal{L}$ using $\wedge, \vee, \bot$ and $\exists$.*

Definition 7. *A formula in $\mathcal{L}$ is a geometric implication if it is of the form $\forall \bar{x}(\phi(\bar{x}) \to \psi(\bar{x}))$, where ϕ, ψ are geometric formulas in $\mathcal{L}$*

Clearly, a geometric formula ϕ is a special case of a geometric implication, where the antecedent is a conjunction of identity statements and the consequent is ϕ itself. Geometric logic is then often taken to be the logic of geometric implications (Achourioti & van Lambalgen, 2011; Dyckhoff & Negri, 2015). Note that geometric implications can be rewritten into a more manageable form by the following lemma:

Lemma 1. *A geometric implication is equivalent to a conjunction of formulas $\forall \bar{x}(\theta(\bar{x}) \to \psi(\bar{x}))$, where θ is a conjunction of atomic formulas and ψ is geometric.*

Proof. By basic logical manipulation. For example, we can first pull the existential quantifiers in the antecedent out using $\exists \bar{x}\, \phi(\bar{x}) \vee \exists \bar{x}\, \psi(\bar{x})) \equiv \exists \bar{x}\, (\phi(\bar{x}) \vee \psi(\bar{x}))$ and $\exists \bar{x}\, \phi(\bar{x}) \wedge \exists \bar{x}\, \psi(\bar{x})) \equiv \exists \bar{x}\bar{y}\, (\phi(\bar{x}) \wedge \psi[\bar{y}/\bar{x}](\bar{y}))$. Subsequently, we can apply the distributive laws for $\vee$ and $\wedge$ to transform the antecedent into a formula in disjunctive normal form with an existential prefix. Since $(\exists x \psi(x)) \to \phi \equiv \forall \bar{x}(\psi(x) \to \phi)$, and since $A \vee B \to C \equiv (A \to C) \wedge (B \to C)$ we arrive at the desired form. $\square$

[1] When writing $\phi(\bar{x})$, The notation $\bar{x}$ represents a vector containing all free variables in ϕ, although not all variables in $\bar{x}$ must necessarily occur in ϕ for ease of notation

Achourioti and van Lambalgen argue that geometric implications allow for Kant's 'mediate cognition of objects', by the subsumption of general objects (as logical *types*) under concepts and the construction of transitive relations between concepts as conditions (2017). Interestingly, the logic of geometric implications is also *intuitionistic* (Achourioti & van Lambalgen, 2011) (Palmgren, 2002):

Lemma 2. *Let $\Gamma \cup \{\phi\}$ be a finite set of geometric implications such that $\Gamma \vdash \phi$ classically, then there exists an intuitionistic proof of ϕ from Γ.*

From a philosophical point of view, this shows that the logic of geometric implications can be considered 'constructive'. Coquand (2010) has shown the completeness for geometric logic of *dynamical proofs*, which bring out this constructive character nicely. Dynamical proofs proceed by model construction, where existential quantification introduces a witness term 'x' and implication corresponds to inference. This then shows how judgements as geometric implications can be understood as Kantian *licenses for inference*. Furthermore, Achourioti and van Lambalgen have shown that finite conjunctions of geometric implications correspond to the class of formulas which are 'objectively valid' with respect to inverse systems, meaning that if a formula is satisfied on all models of an inverse system it must also be satisfied on its inverse limit (2011). The latter semantics then represents Kant's notion of judgement as "the way to bring given cognitions to the objective unity of apperception" (B142).

Since geometric logic is intuitionistic it adheres to the disjunction property: $\vdash A \vee B$ implies that either $\vdash A$ or $\vdash B$. Hence, given the dynamical proof associated with the geometric implication $\phi \to (\psi \vee \xi)$ (where again implication represents inference), there must exist either an inference for $\phi \to \psi$ or for $\phi \to \xi$. In this sense, we can refrain from using disjunctions in the consequents without restricting the descriptive capacity. This brings us to the final form of geometrical rules to be implemented:

Definition 8. *Given a first-order language $\mathcal{L}$, a formula is positive primitive in $\mathcal{L}$ if it is of the form $\exists \bar{x}(\phi_1(\bar{x}) \wedge \cdots \wedge \phi_n(\bar{x}))$, where the ϕ_i are atomic formulas in $\mathcal{L}$.*

Definition 9. *A Geometric rule is a formula of the form $\forall \bar{x}(\theta(\bar{x}) \to \psi(\bar{x}))$, where θ is a conjunction of atomic formulas and ψ is positive primitive.*

I then define GEOLOG as the logic of geometric rules. Of course, disjunctions in the consequents of implications might be quite useful for the construction of theories that explain sensory sequences. For instance, the sensed atoms $bird(a), flies(a), bird(b), walks(b)$ can be 'explained' by a rule the form $\forall x(bird(x) \to (flies(x) \vee walks(x))$. However, it is not at all clear that an explanation of this form is in agreement with Kant's table of judgements. Disjunctive judgements for Kant take the form of reciprocal interaction that represents division of a concept into mutually exclusive 'spheres' under the category of community (A74/B99). One may further ask whether the full complexity of GEOLOG is needed to represent Kant's table of judgements, and if DATALOG is not already sufficient. Achourioti and van Lambalgen argue that the full complexity of geometric rules is indeed necessary, providing the following interpretation of a hypothetical judgement distilled from the *Prolegomena* (29):

"If x is illuminated by y between time t and time s and $s - t > d$ and the temperature of x at t is v, then there exists a w such that the temperature of x at s is $v + w$ and $v + w > c$." (Achourioti & van Lambalgen, 2011, p.8)

Here, d is a criterion for 'long enough' and c is some threshold value for 'warm'. The objective for the rest of this chapter is then to provide a satisfying implementation for this judgement; an implementation in which existential quantification allows for the subsumption of representations with an intensive degree (i.e. a numerical value) under a universal condition, allowing their embedding in a conceptual structure of rules.

3.3 Datalog with choice

In order to enable the APPERCEPTION ENGINE to construct rules in geometric logic, extending the universally quantified implications of DATALOG$^{\ni^-}$, I add a 'choice' predicate to the object language. The predicate 'choice($\bar{X}, \bar{Y}$)' represents a non-deterministic choice between the possible substitutions for $\bar{Y}$ for each substitution for $\bar{X}$. As a general example, the following rule represents the implication $\forall \bar{x}\bar{w}\,(q(\bar{x}, \bar{w}) \rightarrow \exists \bar{y}(p(\bar{x}, \bar{y}))$:

```
P(X̄, Ȳ) ← Q(X̄, W̄), choose(X̄, Ȳ).
```

This rule can then be defined in a language with negation (DATALOG$^{\neg}$) as follows:

```
P(X̄,Ȳ) ← Q(X̄,W̄), choice(X̄,Ȳ).

choice(X̄,Ȳ) ← possible_choice(X̄,Ȳ),¬diff_choice(X̄,Ȳ).

diff_choice(X̄,Ȳ) ← choice(X̄,Ȳ′), possible_choice(X̄,Ȳ), Ȳ′! = Ȳ.
```

Program 3.1: Definition of the choice predicate in DATALOG$^{\neg}$

Here the 'possible_choice' predicate defines the domain of possible substitutions for $\bar{Y}$ given $\bar{X}$. Note that existential quantification requires the system to find a witness term for $\bar{y}$, analogous to the structure of dynamical proofs. In what follows, I will elaborate on the representation of *geometric rules* more formally. The experienced reader will however have noticed that this definition of choice makes use of non-stratified (i.e., recursively applied) negation. Hence, before I explain how the choice predicate is applied, I must properly delineate the intended semantics for a DATALOG$^{\neg}$ program.

3.3.1 Semantics for unstratified negation

Rules in DATALOG$^{\neg}$ have the following form:

$$a_0 \leftarrow a_1, \ldots, a_m, \neg a_{m+1}, \ldots \neg a_n.$$

where the a_i are function-free atoms as before, and rules are still restricted to be safe: variables in the head also occur in the body. In general, a distinction is made between *fixed-point* and *model-based* semantics for DATALOG$^{\neg}$ (Abiteboul, Hull, & Vianu, 1995). The first interprets atoms as either true or false, while the latter often allows for the intermediate valuation

$\frac{1}{2}$. Fixed-point semantics correspond to the stable-model semantics in ASP. Given an interpretation I as a set of positive atoms, the reduct P^I is generated by substituting all negative atoms $\neg\alpha$ in bodies with 0 if $\alpha \in I$ and 1 otherwise. The minimal model of the resulting positive logic program can then be computed as the fixed point of the consequence operator $T_{P^I}(\emptyset)$ as defined before, and I is a model for P if and only if it is equal to $T^{\infty}_{P^I}(\emptyset)$.

Fixed point semantics display downwards monotonic behavior (if $I \subseteq J$ then $T^{\infty}_{P^J}(\emptyset) \subseteq T^{\infty}_{P^I}(\emptyset)$), which is considered to be a useful tool in the implementation of non-monotonic reasoning. However, it also has properties that can be considered disadvantages. Programs are not guaranteed to have a model (e.g., $p \leftarrow \neg p.$), but can also have several subset-minimal models (e.g., $p \leftarrow \neg q.\ q \leftarrow \neg p.$ has models $\{p\}, \{q\}$). The consequence operator T_P can no longer be guaranteed to reach a fixed point as was the case for positive logic programs. One solution for these issues is the restriction to DATALOG$^{\neg s}$, which only allows for *stratified negation*. Formally, this requires programs to be divisible into a partition of programs $P = \{P^1, \ldots, P^n\}$, where all rules with the same predicate p in the head are in the same element of the partition, provided by a function $f(p)$ that maps predicates to numbers $1 \leq i \leq n$ and:

- If a predicate p occurs positively in the body of a head with predicate q, then $f(p) \leq f(q)$.
- If a predicate p occurs negatively in the body of a head with predicate q, then $f(p) < f(q)$.

Under this restriction one can find a sequence of models $I^1 \ldots I^n$ where:

$$I^1 = T^{\infty}_{P^1}(\emptyset)$$
$$I^i = I^{i-1} \cup T^{\infty}_{P^{I^{i-1}}|}(\emptyset)$$

Note that we now iterate over fixed points of the T operator, applying each fixed point I^i to redefine the reduced program for the next operator $T_{P^{I^i}}$. The intended model for P is then equal to I^n. Clearly, an atom is only added to I^i if it is true in every interpretation satisfying the implications of P, and I^n is an interpretation for P. Hence I^n is a unique minimal model, and in fact one that is computable in polynomial time with respect to P (Abiteboul et al., 1995).

While an extension of the object language to DATALOG$^{\neg s}$ can thus be implemented with relatively little impact on the structure of the AE, it is not sufficient to define non-deterministic choice. One may then wonder whether an alternative *model-based* semantics would allow the use of unstratified negation while guaranteeing the existence of models for DATALOG$^{\neg}$ programs. Under model-based semantics, literals are generally valuated as a value in $\{0, \frac{1}{2}, 1\}$, so that partial models are allowed in which not all atoms are decided (Abiteboul et al., 1995). Accordingly, the reduct of a program P with respect to a model M is computed by substituting all negated atoms $\neg\alpha$ by $1 - v_M(\alpha)$. Such a reduced program again has a unique minimal model that can be computed using a consequence operator T_{P^M}, but now this is a three-valued model, where the valuation $v(\alpha)$ in model $T_P(M)$ is:

$$\begin{cases} \mathbf{1} \text{ if there is a rule } r \text{ with head}(r) = \alpha \text{ and } \forall \beta \in \text{body}(r) : v_M(\beta) = 1 \\ \mathbf{0} \text{ if for all rules } r \text{ with head}(r) = \alpha \text{ it holds that } \exists \beta \in \text{body}(r) : v_M(\beta) = 0 \\ \quad \text{hence also if there are no rules with } \alpha \text{ as head} \\ \mathbf{\frac{1}{2}} \text{ otherwise} \end{cases}$$

Note that a valuation of $\frac{1}{2}$ here then corresponds to Kleenee's system in the sense that atoms valuated at $\frac{1}{2}$ may stay $\frac{1}{2}$ or become 0 or 1 by applying the consequence operator T. Now again, models of a program P are fixed-points, i.e. $M = T_{P^M}(\emptyset)$. Interestingly, under these semantics every program P has a unique minimal model called the *well-founded model* P^{wf} (Abiteboul et al., 1995, p.390-394). It is minimal in the sense that its positively and negatively valuated atoms are a subset of the positive and negative atoms for all other models of P. For stratified programs the minimal model is in fact a *total* model, no atoms are interpreted as $\frac{1}{2}$. Furthermore, the well-founded model can be computed in polynomial time with respect to P (Greco, 1998). As for stratified programs we can construct a sequence of total models $I^0 \dots I^n$, but now $I^0 = \bot$ (i.e. all atoms are false), and $I^i = T^{\infty}_{P^{I^{i-1}}}(\emptyset)$. In this sequence then I^0 is an overestimation of the negative facts, so that I^1 is an overestimation of the positive facts, so that I^2 is an overestimation of the negative facts etc.. The process either oscillates towards a total model, or converges to an alternating sequence $I^*, I^{**}, I^*, I^{**} \dots$. In the latter case, atoms that are decided differently by I^* and I^{**} are valuated as $\frac{1}{2}$. The reader may read (Abiteboul et al., 1995, p.390-392) for a more complete explanation.

Thus, if one applies model-based semantics and takes the well-founded model as the intended interpretation this has some model-theoretic advantages. Programs can use negation freely, while a unique minimal model can be computed in polynomial time. However, this unfortunately does not fit our desiderata. One can easily see that if we include choice predicates, then the atoms with predicates 'diff_choice' and 'choice' are normally interpreted as $\frac{1}{2}$ in the well-founded model. The choice predicate represents a non-deterministic choice, while the well-founded model only includes those positive and negative valuations that are true in *every model*. In fact, only if there is exactly one $\bar{Y}$ for a given substitution of $\bar{X}$ such that 'possible_choice($\bar{X}, \bar{Y}$)' holds will the well-founded model valuate diff_choice($\bar{X}, \bar{Y}$) as 0 and choice($\bar{X}, \bar{Y}$) as 1.

Since the only model of which the existence for DATALOG$^{\neg}$ programs is guaranteed is the well-founded model, which does not fit our intended use, one may wonder whether we should content ourselves with the occurrence of programs without a model, even before the introduction of constraints. If we however restrict our syntactical structure to provide just the needed amount of expressivitiy, but no more, this need not be the case. Besides the well-founded model other commonly used partial stable models are (Greco, 1998):

- Total stable models, which do not valuate any atom as $\frac{1}{2}$.
- Maximal stable models, for which there does not exist a distinct stable model which is a strict superset.

- Least-undefined stable models, for which the set of undefined atoms is minimal; there does not exist a distinct stable model for which the set of undefined atoms is a strict subset.

Now the following was shown by Giannotti, Pedreschi, Sacca, and Zaniolo (1991) for the language DATALOG$^{\neg s,c}$, i.e. the language that includes choice predicates, but where all other negations are stratified.

Lemma 3. *Let P be a program in* DATALOG$^{\neg s,c}$*. If none of the rules in which a choice-predicate occurs are recursive, then P has at least one total stable model. Furthermore, one of these models can be computed in polynomial time.*

Hence, in this case the maximal and least-undefined stable models must also be total stable models. Clearly, total stable models are also 'choice models', in the sense that they satisfy a constraint of functional dependency $\bar{X} \rightarrow \bar{Y}$ for every choice predicate 'choice($\bar{X}, \bar{Y}$).

Total stable model semantics (or equivalently maximal or least-undefined stable model semantics) is thus a suitable choice of semantics for our system. It also fits with our desired choice of implementation since it corresponds to fixed-point semantics and thus also to stable model semantics in ASP in the following way:

Theorem 1. *Given a program P in* DATALOG$^{\neg}$*, there exists a function-free ASP program P', constructed from P by replacing each $\leftarrow$ by $:-$ and each $\neg$ by 'not', so that there exists a bijection f between total models M for P and interpretations I' for P', such that for each total model M, $f(M) = \{\alpha : v_M(\alpha) = 1\}$ and for each interpretation I' for P',* $v_{f(I')}(\alpha) = 1$ if $\alpha \in I'$ and $v_{f(I')}(\alpha) = 0$ otherwise .

Proof. Clearly, P' is an ASP program and f is a bijection between total valuations and interpretations. It remains to be shown that M is a model for P iff $f(M)$ is a model for P'. The reducts P_M and $P'_{f(M)}$ correspond to the same positive logic program, since M is total and since in P_M all literals $\neg\alpha$ are substituted with 1 if $v_M(\alpha) = 0$ and with 0 otherwise and in $P'_{f(M)}$ all literals not α are substituted with 1 if $\alpha \notin f(M)$ and with 0 otherwise. Since $v_M(l) = 1$ iff $l \in f(M)$ for each positive litreal l, we thus have $M = T^{\infty}_{P_M}(\bot)$ iff $f(M) = T^{\infty}_{P'_{f(M)}}(\emptyset)$. □

3.3.2 Constraints and Causality

Naturally, not all off the above results hold for DATALOG$^{\ni,\neg s,c}$, which includes the constraints and causal rules introduced by Evans (2020). Causal rules give rise to infinite models, so that these models can clearly no longer be computed in polynomial time. However, if we restrict the interpretation to a predefined set of times $\{1, \ldots, n\}$, as is sufficient to assess whether an *apperception task* has been solved and as is done in the AE's implementation, we can substitute all causal rules of the form:

$$p(\bar{X}) \in q_1(\bar{X}), \ldots q_n(\bar{X}).$$

where the q_i are positive or negative atomic formulas, with the equivalent set of function-free rules for $1 \leq t < n$:

$$p(\bar{X}, t+1) \leftarrow q_1(\bar{X}, t), \ldots q_n(\bar{X}, t).$$

Then given the maximum time n, every DATALOG$^{\ni,\neg s,c}$ P program without constraints is equivalent to a DATALOG$^{\neg s,c}$ program P_n constructed by the procedure above, and thus by lemma 3 has a model that can be computed in polynomial time if its choice rules are not recursive. The same holds if we further include the 'frame axiom' without yet making use of constraints, since this axiom can then be translated into rules of the following form for each variable atom $p(\bar{X})$:

$$p(\bar{X}, t+1) \leftarrow p(\bar{X}, t)$$

If we allow for constraints, there are clearly DATALOG$^{\ni}$ programs that do not have a model, such as the following example, where a is in the domain of X:

```
p(a).
q(X) ← p(X).
p(X) XOR (X).
```

Furthermore, using XOR constraints we can easily construct a program $P(\phi)$ of size polynomial in the size of propositional logic formula ϕ, such that the models of $P(\phi)$ are satisfying assignments for ϕ, by iteratively splitting subformulas $\psi = \psi_1 \circ \psi_2$ or $\psi = \neg\psi_1$ of ϕ (starting with ϕ itself) into smaller subformula(s) ψ_1, ψ_2 and adding the following rules to $P(\phi)$ according to the main connective c of ψ, I assume ψ is in CNF since this is achievable in polynomial time:

$$\begin{aligned} c = \wedge : \psi(1) &\leftarrow \psi_1(1), \psi_2(1). \\ \psi(0) &\leftarrow \psi_1(0). \\ \psi(0) &\leftarrow \psi_2(0). \end{aligned}$$

$$\begin{aligned} c = \vee : \psi(1) &\leftarrow \psi_1(1). \\ \psi(1) &\leftarrow \psi_2(1). \\ \psi(0) &\leftarrow \psi_1(0), \psi_2(0). \end{aligned}$$

$$\begin{aligned} c = \neg : \psi(1) &\leftarrow \psi_1(0) \\ \psi(0) &\leftarrow \psi_1(1) \end{aligned}$$

We can then add the following constraint for all subformulas ψ of ϕ (polynomially many in its size).

$$\psi(1) \; XOR \; \psi(0)$$

The models for $P(\phi)$ then clearly correspond to satisfying assignments for ϕ, showing that finding a model for a DATALOG$^{\ni,\neg s,c}$ program is in general NP hard. Since DATALOG$^{\ni,\neg s,c}$ restricted to a fixed set of times can be expressed in ASP without disjunctions in the head, for which model-existence is NP-complete, this shows that finding a model for a DATALOG$^{\exists,\neg s,c}$ program is also NP-complete.

Finally, DATALOG$^{\exists,\neg s,c}$ programs are clearly equivalent to a subset of DATALOG$^{\neg}$ programs, so that theorem 1 applies: programs in DATALOG$^{\exists,\neg s,c}$ can be translated into equivalent ASP programs under the total-model or fixed-point semantics.

3.4 Implementation

3.4.1 Choice

A complete overview of the DATALOG$^{\exists}$ interpreter developed by Evans is provided in *Kant's Cognitive Architecture* (Evans, 2020, p.58-63). I take this as given and do not repeat the general structure here. Instead, I emphasize the changes made to represent rules in geometric logic using choice predicates.

Recall that geometric rules are of the form $\phi_i = \forall \bar{x}\bar{w}\,(\theta_i(\bar{x},\bar{w}) \rightarrow \psi_i(\bar{x}))$ where θ_i is a conjunction of atomic formulas and ψ_i is an existentially quantified conjunction of atomic formulas $\exists \bar{y}\,\xi_i(\bar{x},\bar{y})$. The universally quantified variables that only occur in the antecedent are identified with $\bar{w}$. The variables in $\bar{x}$ need not necessarily occur in θ_i. They might occur in both the antecedent and the consequent or only in the consequent. Such a ϕ_i is then represented by the following rule in DATALOG$^{\exists,\neg s,c}$, where r is a numerical index for the rule and T is a variable with the domain of times. $(\overline{X,T,r})$ represents the variables in $\bar{X}$ and T as well as the number r. Note that this is only for ease of notation however, since the complex term $(\overline{X,T,r})$ would itself not be allowed in function-free DATALOG:

```
ξi(X̄,Ȳ) ← θi(X̄,W̄), choose((X,T,r),Ȳ).
```

Note that we enforce a non-deterministic choice over the domain of $\bar{Y}$ for each pairing of a substitution for $\bar{X}$, time t and index r. Hence, for two distinct rules with the same variables $\phi_1 = \forall \bar{x}(\theta_1(\bar{x}) \rightarrow \exists \bar{y}\xi_1(\bar{x},\bar{y}))$ and $\phi_2 = \forall \bar{x}(\theta_2(\bar{x}) \rightarrow \exists \bar{y}\xi_2(\bar{x},\bar{y}))$ the choice need not be the same, nor need the choice be the same for two applications of the same rule at different times. This then gives the following clause in DATALOG$^{\neg}$:

```
ξi(X̄,Ȳ) ← θi(X̄,W̄), choice((X,T,r),Ȳ).

choice((X,T,r),Ȳ) ← possible_choice((X,T,r),Ȳ),
    ¬diff_choice ((X,T,r),Ȳ).

diff_choice((X,T,r),Ȳ) ← choice((X,T,r),Ȳ'),  possible_choice((X,T,r),Ȳ),  Y'!=Y.
```

Program 3.2: DATALOG$^{\neg}$ program corresponding to ϕ_i

In the interpreter in the meta-language ASP, I define a new type of rules, apart from the 'arrow rules' and 'causal rules' in the original system. I name them 'causal judgements' since I apply the structure of geometric logic for construction of causal judgements, but extensions to other judgements can easily be made. Now in the original system, the consequent of causal rules is derived by the following ASP clause:

```
holds(GC, T+1) :-
    rule_head_causes(R, VC),
    eval_body(R, Subs, T),
    ground_atom(VC, GC, Subs),
    is_time(T+1).
```

Program 3.3: Derivation of the head of causal rules for the original AE

Here, 'GC' is a representation of the ground atom in the head where the first argument represents a predicate symbol in the object-language and the other arguments represent constants (e.g., $s(c\_p, obj\_a)$ or $s2(c\_p, obj\_a, obj\_b)$). R is a number referring to the rule, VC is a representation of an *unground* atom (e.g., $s(c\_p, var\_x)$ or $s2(c\_p, var\_x, var\_y)$), and 'Subs' represents a substitution that substitutes the variables in VC with the corresponding constants in GC. 'eval_body' is true if and only if the ground atoms produced by applying 'Subs' to the body of 'R' hold at time T. I substitute this clause with the following:

```
holds(GC, T+1) :-
    causal_judgement_head(R, VC),
    eval_causal_judgement_body(R, Subs, T),
    ground_atom(VC, GC, Subs),
    is_time(T+1),
    rule_choice(R, T, Subs),
    not is_var_permanent(VC).
```

Program 3.4: Derivation of the head of causal rules for the system with existential quantification

The final line is prompted by a technical difference with the original system. In the original AE, 'permanent' atomic formulas (such as those determining relations of arithmetic) can never be in the head of a rule by construction, since formulating rules with known information in the consequent would be a waste of resources. However, since I allow for *multiple atoms* in the head of a rule, it might often be sensible to construct rules in which some of these atoms represent known information. This then restricts the domain of choice for the consequent $\exists \bar{y}\, \xi_i(\bar{y})$, since only substitutions that satisfy the 'permanent' atoms can possibly satisfy ξ_i. However, these permanent atoms need not be derived themselves so that they are excluded in the clause above for clarity and efficiency.

The second-to last line represents the choice predicate and is defined by the following clauses:

```
rule_choice(R, T, Subs) :-
    possible_subs(R, Subs), is_time(T),
    not diff_choice(R, T, Subs).

diff_choice(R, T, Diff) :-
    rule_choice(R, T, Subs), possible_subs(R, Diff),
    rule_tochoose(R, V1), subs(Subs, V1, O1),
    subs(Diff, V1, O2), O1 != O2,
    subs(Diff, V2, O3) : rule_ground(R, V2), subs(Subs, V2, O3).

possible_subs(R, Subs) :-
    rule_subs(R, Subs), is_causal_judgement(R), use_rule(R),
```

```
    permanent(GH) : causal_judgement_head(R, Head),
    ground_atom(Head, GH, Subs),
    is_var_permanent(Head).

:- eval_causal_judgement_body(R, Subs, T),
   not possible_subs(R, Diff) : same_ground(R, Subs, Diff).
```

Program 3.5: Generation of choices for existential quantication

Here the 'rule_tochoose' and 'rule_ground' predicates identify the variables in $\bar{y}$ and $\bar{x}$ respectively in the corresponding ϕ_i. The clauses defining the 'rule_choice' and 'diff_choice' predicates are direct analogues of the second and third DATALOG$^{\exists,\neg s,c}$ rules in program 3.2. Note that $Y \neq Y'$ (while $X = X$) here means that 'Subs' and 'Diff' are two different choices for the 'to_choose' variables corresponding to the same choice for the 'ground' variables. The third clause defines the domain of possible choices, where possible choices are constrained by the permanent conjuncts in the head of a rule, as explained above. The final constraint ensures that if the body of a rule is satisfied for some substitution, there must be at least one possible extension of this substitution to the variables in the head that does not conflict with the permanent information, so that a choice must be made.

Note that these clauses indeed ensure that the formula $\phi_i = \forall \bar{x}\bar{w}\, \forall 1 \leq t \leq n\, (\theta_i(\bar{x}, \bar{w}, t) \rightarrow \exists \bar{y}\, \xi_i(\bar{x}, \bar{y}, t+1))$ is satisfied on all models. If a model satisfies the clauses above, we can partition all substitutions for R in equivalence classes so that substitutions in the same class assign the same values to variables in 'rule_ground' (representing $\bar{x}$), and so that at least one chosen substitution in each equivalence class where 'eval_causal_judgement_body' is true (i.e., $\theta_i(\bar{x})$ is satisfied) also makes 'holds(GC,T+1)' true for all ground instances GC of unground atoms in the head of R. The latter is ensured by the final constraint in Program 2. Thus, whenever a substitution for $\bar{x}, \bar{w}$ makes θ_i true, it must be extendable to a substitution that also makes ξ_i true. Conversely, if a model satisfies ϕ_i, then every substitution σ for the variables in ϕ_i produces a true ground instance of ϕ_i. Hence, whenever a substitution for $\bar{x}, \bar{w}$ makes θ_i true, it must be extendable to a substitution that also makes ξ_i true. Hence, of all substitutions that make 'eval_causal_judgement_body' true for some T and the R corresponding to ϕ_i, there must be at least one that also makes 'holds(GC,T+1)' true for all 'GC' representing a ground conjunct of ξ_i.

It is of importance for efficiency that the variables in $\bar{w}$, i.e. variables that only occur in the antecedent of a ϕ_i, are omitted from the ground. If this were not the case, distinct choices for $\bar{y}$ would be generated for each suitable choice of $\bar{w}$, while in fact the generation of a single choice for some substitution of $\bar{x}$ (that can be extended to satisfy θ_i) is sufficient to satisfy ϕ_i. Note further that it is correct that the universally quantified variables occurring only in the consequent are part of the 'ground'. ϕ_i enforces that for any substitution satisfying its antecedent and *all substitutions for the universally quantified variables in the consequent* there exists a suitable choice for $\bar{y}$. If we would not place such variables in the ground, some models would not satisfy the constraint represented by a ϕ_i, but would satisfy the corresponding ASP clauses. To see this, consider the formula $\phi_i = \forall x\, y\, (p(x) \rightarrow \exists z\, r(x, y, z))$. The model $M = \{p(a), r(a, b, c)\}$ then clearly does not satisfy ϕ_i, but does satisfy the constraint that for each substitution for x satisfying $p(x)$ there is a suitable extension.

3.4.2 Rule generation

In the generation of rules from a template I make a few simple adjustments to the mechanism of the AE. Firstly, the number of possible heads for a 'causal judgement' is now allowed to be larger than 1. Secondly, atoms representing 'permanent' information are now allowed to be in the head of rules. Thirdly, I generate a division of the variables in a rule in three categories: "ground', to_choose', and 'independent', representing $\bar{x}, \bar{y}$ and $\bar{w}$ in the corresponding formula. This is done by generating facts of the following form using Haskell:

```
1 { rule_ground(R,V), rule_tochoose(R,V), rule_independent(R,V) } 1 :-
    is_causal_judgement(R), use_rule(R),
    rule_var_group(R,VG), contains_var(VG,V).
```

Program 3.6: Choice of variable distribution over a causal judgement

Now in order to ensure that the chosen division indeed corresponds to the variable groups $\bar{x}, \bar{y}, \bar{w}$ in ϕ_i, I formulate several constraints: 'ground' variables must occur in the head and can occur in the body, 'tochoose' variables must occur only in the head, and 'independent' variables may not occur in the head and can occur in the body, but can also simply be a redundant part of the variable group:

```
violation(ground_notin_head(R, V)) :-
    rule_ground(R, V),
    not var_in_caus_judg_head(V, R).

violation(tochoose_notin_head(R, V)) :-
    rule_tochoose(R, V),
    not var_in_caus_judg_head(R, V).

violation(tochoose_in_body(R, V)) :-
    rule_tochoose(R, V),
    var_in_caus_judg_body(V, R).

violation(independent_in_head(R, V)) :-
    rule_independent(R, V),
    var_in_caus_judg_head(V, R).

:- violation(_).
```

Program 3.7: Constraints on the variable distributions for causal judgements

I further delete the 'safeness' constraint applied in the original system, since geometric rules can have both universally and existentially quantified variables that occur only in the head. Formally, this does not imply that our object-language transcends DATALOG$^{\exists,\neg s,c}$, since all variables have been assigned a predefined domain by the type-signature in the template. Hence, we may simply add clauses of the form 'dom_$x(X)$ in the body of a rule for every of its 'unsafe' variables X to produce a safe program with the same interpretation. Practically, I simply remove the safeness constraint for causal judgements from the program.

3.4.3 Interlude: the trouble with trace multiplicity

We have noted in chapter 1 that the *apperception task* that is to be solved by the AE comes down to the formulation of a conceptual theory that satisfies several unity conditions and of which the trace covers the provided sensory sequence. We have also noted that positive logic programs, such as the restrictions to finite time of constraint-free DATALOG$^{\ni}$ programs, have a unique minimal model corresponding to a unique total model. Hence, if a DATALOG$^{\ni}$ program *with* constraints has a model, it must also have unique minimal model. The relations between conceptual structures and sensory sequences is then immediately clear: conceptual structures generate a single trace and the apperception task is completed if and only if this trace covers the input sequence.

With the introduction of existentially quantified variables in the head the situation becomes altogether different. A rule now ensures that the consequent is true for *some* choice of substitution. Each choice generates a distinct ground consequent, so that each set of choices generates its own trace and theories have *multiple traces*. In order to maintain the existential interpretation of the choice predicate, I insist that *at least one* of these traces must explain the input sensory sequence. For instance, the following initial conditions and rule explain (along with a suitable domain for the variable Y, say {alice, bob, charlie}) every sensory sequence in which each child has at most one parent such as $\{parent(bob, charlie), parent(bob, alice)\}$:

```
child(alice).
child(charlie).
parent(Y,X) ← child(X), choose(X,Y).
```

We would not insist that *all traces* must cover the sensory sequence, or this theory would only explain sequences without any parents. The extension of the 'parent' relation in a sequence explained by this theory would have to be subsumed by both $\{parent(alice, charlie), parent(alice, alice)\}$ and $\{parent(charlie, charlie), parent(charlie, alice)\}$ so that it must be empty.

This approach however also has some disadvantages. Firstly, checking whether a theory explains a sequence in the worst case now requires computing all of its traces. While these are exponentially many in the number of choices, this does not affect the complexity of the underlying problem of finding a model in the meta-language. The general problem in the meta language is still to find a model for an ASP program with disjunctions in the head and optimization statements, which is Δ_3^P-complete (Gebser et al., 2012). Secondly, merely requiring that *some* trace explains a sensory sequence might be an awfully easy thing to ask of the AE. For example, consider the following sequence:

$$S = (\{p(0)\}, \{p(1)\}, \{p(2)\}, \{p(3)\}, \{p(4)\}, \{p(5)\}\})$$

$p(0)$ holds at time 1, after which the time steps produce a counting sequence. We would like the AE to pick up on this structure. However, if we merely require subsumption for one of the generated traces, the AE produces the following minimal theory:

$$R = \{\exists x p(x).\}$$

Here, the domain of x is the natural numbers. Our system thus proudly claims: "there exists a number that is p, Problem solved, and with a very parsimonious theory indeed!". Clearly, any sequences of numbers that are 'p' is explained by this rule. We do not want to let the AE off the hook so easily. In the remainder of this chapter I thus explain an implemented solution for this issue that simultaneously addresses the issue of trace explosion: we require the AE to find the model that fits the sensory sequence *as closely as possible*. Closeness here refers to the number of different ground atoms generated by a theory: more precise theories should always be preferred over more general theories, and the quality of explanation must be maximized before the size of the theory is minimized.

3.4.4 Choice minimization

We want the AE to construct those theories that are *generalizable*, in the sense that omitted sensory atoms (i.e., new data points) are predicted with high accuracy. In a machine learning context, performance is usually based upon the prediction error on a test-set. Of course, we cannot let the AE minimize the error on omitted atoms directly, since this would come down to 'peeking' at the test set. However, we also do not want to burden the system with the task of minimizing the possibilities for its whole sensory sequence. This is not only computationally inefficient, but it might even cause over-fitting: explaining the sensory sequence so precisely that the theory becomes less generalizable. I thus provide the system with a 'validation set', as is common in machine learning. Besides the sensed 'training' atoms, and the 'test' atoms on which performance is evaluated, the atoms in the validation set have the sole purpose of maximizing the generalizability of the constructed theory.

When a validation atom follows from a rule without existential quantification it is derived with certainty. The number of choice possibilities for the predicted atom is thus 1:

```
possible_predictions(GA, T, 1) :-
    validation(GA, T),
    rule_arrow_head(R, VA),
    ground_atom(VA, GA, Subs),
    eval_body(R, Subs, T).

possible_predictions(GA, T, 1) :-
    validation(GA, T),
    rule_arrow_head(R, VA),
    ground_atom(VA, GA, Subs),
    eval_body(R, Subs, T),
    not head_in_body(R, Subs).

head_in_body(R, Subs) :-
    causal_judgement_head(R, VC),
    rule_body(R, VA),
    ground_atom(VC, GC, Subs),
    ground_atom(VA, GC, Subs).
```

Program 3.8: Determining when an atom is derived with certainty

The 'head_in_body' predicate ensures that a rule does not count as an explanation if it already presupposes the truth of the head. This prevents the construction of theories with 'fake generalizability'. For example, the series of numbers from the interlude might be explained by the following theory, in which the first rule generates a trace, and all other rules act to 'derive' an atom with certainty, but can only do so under presupposition of its truth:

$$R = \begin{Bmatrix} \exists x\, p(x). & \\ first\_number(x) \land p(x) \land first\_number(y) & \to p(y). \\ first\_number(x) \land succ(x, x\_1) \land p(x\_1) & \\ \land\, succ(x, y) & \to p(y). \\ \ldots & \end{Bmatrix}$$

this example then also shows why the 'head_in_body' predicate is necessary, even though the original system already prevents rules from having the same variable atom in the head and body, since different variable atoms can produce the same ground atom.

Now for each validation atom that is one of several possibilities provided by a 'causal judgement', I identify the substitution that produces this atom and record the different choices that could have been made for the same 'ground':

```
different_option(R, Subs, Val, T, Notval) :-
    validation_explanation(R, Subs, Val, T),
    causal_judgement_head(R, VC),
    same_ground(R, Subs, Diff),
    eval_body(R,Diff,T-1),
    ground_atom(VC, Val, Subs),
    ground_atom(VC, Notval, Diff),
    Val != Notval.

validation_explanation(R, Subs, Val, T) :-
    validation(Val, T),
    causal_judgement_head(R, VC),
    ground_atom(VC, Val, Subs),
    eval_causal_judgement_body(R, Subs, T-1),
    possible_subs(R, Subs).

same_ground(R, Subs1, Subs2) :-
    is_causal_judgement(R),
    use_rule(R),
    possible_subs(R, Subs1),
    possible_subs(R, Subs2),
    subs(Subs2, V, O) :
    rule_ground(R, V), subs(Subs1, V, O).

possible_predictions(Val, T, P) :-
    validation_explanation(R, Subs, Val, T),
    P = #count {Notval : different_option(R, Subs, Val, T, Notval)}.
```

Program 3.9: Counting alternative choices for validation atoms

Here the 'possible_subs' predicate ensures that only substitutions that agree with the 'permanent' background knowledge are taken into considerations. For example, if the rule is of the form $p(x) \to \exists y(y > x \wedge r(x, y))$, then if $x = 2$, substitutions that assign $0, 1$ or 2 to y do not count as possible choices. This predicate was also used to delimit the domain of possible choices in Program 3.5. Note that one 'validation atom' can be predicted by different rules, or even different substitutions for the same rule, so that there can be several numbers of 'possible predictions'. I select the most precise explanation as the minimum over all explanations in a theory, and then select the program that minimizes these minima:

```
minimum_number_of_choices(GA, T, C) :-
    validation(GA, T),
    C = #min { P : possible_predictions(GA, T, P)}.

#minimize {C @ 3, GA, T : minimum_number_of_choices(GA, T, C), C != #sup}.

:- validation(GA, T), not possible_predictions(GA, T, _).
```

Program 3.10: Minimizing the number of choices for validation atoms

Given the sensory sequence from the interlude, the AE now produces the following two equivalent theories of equal cost:

$$\theta_1 \quad : \quad \begin{array}{l} I = \{p(0)\} \\ R = \{p(x) \ni\!\!\!- \exists y(succ(x, y) \wedge p(y))\} \end{array} \qquad \theta_2 \quad : \quad \begin{array}{l} I = \{p(0)\} \\ R = \{p(x) \wedge succ(x, y) \ni\!\!\!- p(y)\} \end{array}$$

3.5 Example behavior

I provide the AE with a sensory sequence that can be explained using the example causal judgement from section 2:

"If x is illuminated by y between time t and time s and $s - t > d$ and the temperature of x at t is v, then there exists a w such that the temperature of x at s is $v + w$ and $v + w > c$." (Achourioti & van Lambalgen, 2011, p.8)

In the system that is described in this chapter, time is a sequence of natural numbers, and causal rules operate from time t to time $t + 1$. Hence, $s - t$ is always 1. In the next chapter I put forward an approach to resolve this. Here, the main point of interest is however the application of geometric logic to find structure in 'vague' sequences that can be explained by 'assertion under a universal condition': the sequence of individual numbers is subsumed universal rules through the condition of being larger than some threshold value.

Recall that for the original AE, templates are of the form $(\phi, N_{\to}, N_{\ni\!\!\!-}, N_B)$, where ϕ is a type signature and the other three arguments are natural numbers representing the number of static rules, causal rules, and maximum body atoms. In our extended setting, templates are now instead defined as $(\phi, N_{\to}, N_{\ni\!\!\!-}, N_{CJ}, N_B, N_H)$, where the two added arguments represent the maximum numbers of causal judgements and head atoms.

Example 2. I provide the AE with the following sensory sequence:"

$$\begin{aligned}
S_1 &= \{temp(rock, 0), sunlight_on(rock)\}\\
S_2 &= \{temp(rock, 5), not_sunlight_on(rock)\}\\
S_3 &= \{temp(rock, 2), not_sunlight_on(rock)\}\\
S_4 &= \{temp(rock, 1), sunlight_on(rock)\}\\
S_5 &= \{temp(rock, 7), sunlight_on(rock)\}\\
S_6 &= \{temp(rock, 8), not_sunlight_on(rock)\}\\
S_7 &= \{not_sunlight_on(rock)\}\\
S_8 &= \{sunlight_on(rock)\}\\
S_9 &= \{sunlight_on(rock)\}\\
S_{10} &= \{not_sunlight_on(rock)\}
\end{aligned}$$

The sensed atoms sunlight_on(rock) and not sunlight_on(rock) are 'exogenous', meaning that the AE is not asked to predict when the sun shines on the rock. Instead, the task for the AE is to find a theory that explains the temperature of the rock, given the information of when the sun shines on it. I provide the following validation atoms, where the i again represent time steps:

$$\begin{aligned}
V_7 &= \{temp(rock, 3)\}\\
V_8 &= \{temp(rock, 2)\}\\
V_9 &= \{temp(rock, 7)\}\\
V_{10} &= \{temp(rock, 9)\}
\end{aligned}$$

The template from which theory construction is executed is the following:

$$\Phi \quad = \quad \left\{\begin{array}{l} T = \{rock, number\},\\ O = \{i : number; 0 \leq i \leq 9\} \cup \{obj\_rock : rock\}\\ P = \{temp(rock, number), threshold(number),\\ \qquad less(number, number), first\_number(number)\\ \qquad succ(number, number)\}\\ V = \{X : rock, Y : number, Z : number\} \end{array}\right\} \begin{array}{c} N_{\rightarrow} = 0\\ N_{\ni} = 0\\ N_{CJ} = 4\\ N_B = 2\\ N_H = 2 \end{array}$$

As explained before, and as was the case in the original AE, the numerical relations 'succ', 'less' and 'first_number' are provided as permanent background knowledge, together with input restrictions preventing an object from having multiple temperatures at the same time. This is then the theory constructed by the AE:

$$I = \Big\{temp(rock, 0), threshold(4)\Big\}$$

$$
\text{R} = \begin{Bmatrix} sunlight\_on(X) \wedge temp(X,Y) & \ni \exists Z\, less(Y,Z) \wedge temp(X,Z), \\ sunlight\_on(X) \wedge threshold(Y) & \ni \exists Z\, less(Y,Z) \wedge temp(X,Z), \\ not\_sunlight\_on(X) \wedge temp(X,Y) & \ni \exists Z\, less(Z,Y) \wedge temp(X,Z), \\ not\_sunlight\_on(X) \wedge threshold(Y) & \ni \exists Z\, less(Z,Y) \wedge temp(X,Z), \end{Bmatrix}
$$

△

Together with the knowledge that objects cannot have two temperatures, these rules then amount to the example causal judgement. The temperature at time 11 is predicted at 1, but could have been any value below the threshold of 4. If the system would have been given another template, containing an extra variable and allowing for the existence of rules with 3 atoms in both head and body, it could have constructed a single rule of the following form:

$$
sunlight\_on(X) \wedge temp(X,Y) \wedge treshold(Z) \ni \exists W less(Y,W) \wedge less(Z,W) \wedge temp(X,W)
$$

Unfortunately, the structure of the AE is such that working with templates of this form requires a lot of memory. The AE finds the trace of a theory by generating all possible substitutions for each group of variables. If the templates allows for the usage of 3 numerical values in a single rule, of which the domain ranges over 10 numbers, this thus means that 10^3 substitutions have to be taken into account throughout the process of trace construction. A smarter way of working with variable substitutions is needed to solve this issue, but this is not in the scope of this thesis.

Chapter 4

Figurative Synthesis and Time

4.1 Introduction

"Space, represented as an object ... contains more than the mere form of intuition, namely the comprehension of the manifold given in accordance with the form of sensibility in an intuitive representation, so that the *form of intuition* merely gives the manifold, but the *formal intuition* gives unity to the representation. In the Aesthetic I ascribed this unity merely to sensibility, only in order to note that it precedes all concepts, though to be sure it presupposes a synthesis, which does not belong to the senses but through which all concepts of space and time first become possible. For since through it (as the understanding determines sensibility) space or time are first given as intuitions, the unity of this a priori intuition belongs to space and time, and not to the concept of the understanding (B161n)

This notorious footnote at B161 of CPR is a much debated cornerstone of Kant's conception of space and time (e.g., (Falkenstein, 2018)). The *Transcendental Aesthetic* paints a picture of space and time as (singular) forms. Here, space and time are instead themselves considered objects, and associated with "comprehension of the manifold". There are two apparent contradictions springing from this dichotomy that must be addressed before we can try to implement Kantian space and time in a computational system. Firstly, it seems contradictory that time and space are both objects and non-objects. Kant is quite explicit in stating that "form of intuition, ... is in itself not an object, but the mere formal condition of one" (A291). Secondly, Kant's claim that the unity of space and time "precedes all concepts" seems difficult to place in his general framework. We have seen that in the Kantian system all synthesis is grounded in the understanding, as Kant clearly claims in B161: "all synthesis, through which even perception itself becomes possible, stands under the categories". How can synthesis simultaneously precede all concepts and be determined by the understanding?

The first paradox is addressed two sections earlier, although indirectly. After having identified figurative synthesis with the *transcendental synthesis of the imagination*, and associating it with *productive imagination* Kant states the following:

"Apperception and its synthetic unity is so far from being the same as the inner sense that the former, rather, as the source of all combination, applies, prior to all sensible intuition of objects in general, to the manifold of intuitions in general, under the name of the categories; inner sense, on the contrary, contains the mere form of intuition, but without combination of the manifold in it, and thus it does not yet contain any determinate intuition at all, which is possible only through the consciousness of the determination of the manifold through the transcendental action of the imagination (synthetic influence of the understanding on inner sense)" (B154).

Formal intuitions are thus the source of combination, applying to a manifold of intuition *a priori*, while the form of intuition does not yet contain any combination or even any *determinate intuition*. While space and time as objects represent the spontaneity of the understanding, as "condition under which all objects of our (human) intuition must necessarily stand" (B151), space and time as forms thus represent the unactualized structure with which all synthesis of apprehension must be in agreement. In this sense space and time affect our experience both as unified objects and forms.

An answer to the second paradox requires a more thorough understanding of the *figurative* synthesis or *transcendental synthesis of the imagination*; a process that was identified as missing in the AE in chapter 2, and that brings forth space and time as *formal intuitions*. In the A-deduction, Kant associates this process with the pure *syntheses of apprehension and reproduction*, which are themselves inseparably combined. Both association and reproduction are necessary to "draw a line in thought" or "think of the time from one noon to the next" (A102). In the B-deduction, he further defines the *synthesis of apprehension* as "the composition of the manifold in an empirical intuition, through which perception ... becomes possible" (B160), and notes subsequently that it must be in agreement with the *forms of intuition*, space and time, which as *formal intuitions* condition this apprehension (B161). The pure figurative synthesis and the unity of space and time as objects thus underlie the possibility of empirical apprehension. Now being unified objects, space and time must in turn stand under the *transcendental unity of apperception* and hence the categories. We thus see that *sensibility* as a capacity for representations that can be unified under concepts (i.e., related to objects), must be receptive to affections *from outside*, as well as *from inside* through the figurative synthesis as the "source of all combination". Following this interpretation, space and time are then the 'most original' effect of the understanding on sensibility, grounding the apprehension of a manifold that is to be reflected under concepts through the functions of judgement. They thus precede any determinate concepts, and hence also the categories as the universal representations of synthesis under concepts, even though the figurative synthesis is itself a "determination of sensibility by the understanding" (Longuenesse, 2020, p.223).

It is then clear that a formal model of Kant's cognitive architecture must assign a central role to *figurative synthesis*. A pre-conceptual determination of sensibility by the understanding must constitute space and time as pure objects and thereby structure intuition for it to be represented under concepts. The aim of this chapter is to provide a computational implementation of this process. In the second, third and fourth sections, I introduce a mathematical model of temporal synthesis through event structures, following work done by Riccardo Pinosio (2017). The mathematical arguments in these sections are however written with the aim of computational implementation in mind, so that mathematical analysis occasionally makes way for evaluation of the proposed computational system within the mathematical framework. In the fifth and sixth sections I then properly describe the mechanics of this system which I name the Figurative Apperception Engine (FAE): a domain-general system that makes sense of perception in a manner that takes figurative synthesis and the formal intuition of time seriously. In the next chapter I extend this approach to space.

4.2 Temporal event structures

4.2.1 Structure of the form of intuition

Before formally representing time as *formal intuition* and *form of intuition*, I recall the structural desiderata that follow from the *Critique*. In the second chapter, I emphasized that time for Kant is infinitely divisible as well as essentially singular, so that "every determinate magnitude of time is only possible through limitation of a single time grounding it" (B48). I also noted that time as a whole (*formal intuition*) grounds succession and simultaneity as relations among perceptions in time. Now these relations can themselves be associated with time as *form of intuition*, as Kant explains in the transcendental aesthetic:

"the time in which we place these representations, which itself precedes the consciousness of them in experience and grounds the way in which we place them in mind as a formal condition, already contains relation of succession and simultaneity, and of that which is simultaneous with succession (of that which persists). Now that which, as representation, can precede any act of thinking something is intuition and, if it contains nothing but relations, it is the form of intuition, which, since it does not represent anything except insofar as something is posited in the mind, can be nothing other than the way in which the mind is affected by its own activity" (B68).

Times in time

Associating time as *form of intuition* with relations of succession and simultaneity, one may ask what the *relata* should be. In the AE, these are points of time represented by natural numbers. This conflicts with Kant's claim that time is infinitely divisible. If time is constructed bottom-up from its points, an infinite set of points must be represented to generate an infinitely divisible time. Kant rather adhered to a top-down conception of *potential* infinity: times can always be divided into smaller times.

"The property of magnitudes on account of which no part of them is the smallest (no part is simple) is called their continuity. Space and time are quanta continua because no part of them can be given except as enclosed between boundaries (points and instants) thus only in such a way that this part is again a space or a time. Space therefore consists only of spaces, time of times. Points and instants are only boundaries, i.e., mere places of their limitation." (A169)

Since parts of time are then themselves times, the relata of time must be extended limitations of time as a whole.

Succession and moments

While the aim is then to move beyond a representation of times in terms of points, it seems that there must still be some form of succession in the input to constrain the search space, like the natural numbers of *subjective time* do for the original AE. The claim that succession in intuition is essential for its synthesis is not at all opposed to Kant's framework, in which moments in time ground the representation of intuition as manifold:

"Every intuition contains a manifold in itself, which however would not be represented as such if the mind did not distinguish the time in the succession of impressions from one another; for as contained in one moment no representation can ever be anything other than absolute unity" (A99).

We have to be careful however not to mistake this succession for a 'property of the input', independent of the faculty of sensibility. This would not at all fit with Kant's aims expounded in the first chapter. Instead, the succession must itself be grounded in the activity of the subject. The following explanation of Beatrice Longuenesse is clarifying:

"The time mentioned here is not that of a succession of impression we might suppose to be given "in itself", prior to even the act of apprehension. The temporality we are dealing with here is generated by the very act of apprehending the manifold." (Longuenesse, 2020, p.37)

In the upcoming sections I explain how the FIGURATIVE APPERCEPTION ENGINE can actively construct a temporal succession of *moments* in the process of *apprehension* and *reproduction*. Still, the practical question of how we should structure the input to constrain the search space remains difficult to answer conclusively. Ideally, we would like to provide the FAE with a continuous stream of input, guided by motion, "as action of the subject"(B155), but this is not feasible in the current setup. The system does not receive sensation as result of its own motion, and since it is built in ASP we have to provide atomic facts (or atomic sensory determinations as in (Evans, 2020, Ch.5)) as its input. One feasible approach is then to maintain the numerical ordering that characterizes input for the original AE. Maintaining the same input format allows sensible comparison between the old and new system. However, for the FAE this input ordering is then merely a starting point from which a small set of logical relations is derived, and the system actively "runs through and take together this manifoldness" (B155) to construct its own independent representation of moments. Still, the reader may rightfully argue that a numerical ordering is not at all a sensible representation of an unprocessed continuous stream. How should we imagine our artificial agent to count its sensation *before* its reception in sensibility? To address this concern I also allow input to be given as partial *event structure*, a term that will be made clear in the next section. This latter format shows more directly how the FAE can use partial temporal structure in the input to construct its own temporal representation of sensation.

4.2.2 A temporal axiom system

Having established that our temporal form of intuition should be a relational structure between times, one must decide which relations and axioms can be used to ensure that our time adequately represents the form of inner sense from the *Critique*. Importantly, this form must be a 'determination of sensibility by the understanding', so that any *a priori* structure must be in accordance with the categories, thereby grounding the applicability of the categories to inner sense:

"the understanding, as spontaneity, can determine the manifold of given representations in accord with the synthetic unity of apperception, and thus think a priori synthetic unity of the appperception of the manifold of sensible inuition, as the condition under which all objects of our (human) intuition must necessarily stand, through which then the categories, as mere forms of thought, aquire objective reality" (B151)

I here use the system developed by Pinosio as the 'logic of Kant's temporal continuum' (2017). This system makes use of the primitive binary relations O, R_+ and R_-. Here, O represents an overlap of temporal events, and is derived from Kant's description of simultaneity as "existence of the manifold at the same time" (B257). aR_-b holds if and only if event a ends before or simultaneously with event b, and aR_+b holds if and only if a begins after or simultaneously with b. Events are thus compared in terms of both their beginnings and endings. This framework is then rich enough to capture Kant's discussion of causality at B248: a cause can begin before or simultaneous with its effect, but the arising of the effect is always simultaneous with the causality of the cause. From the primitive relations an additional covering relation $\preceq$ is defined. The axiom system is then as follows:

1. Definition of covering:
   $a \preceq b \leftrightarrow aR_-b \wedge aR_+b.$

2. Reflexivity and symmetry of overlap:

   a aOa

   b $aOb \rightarrow bOa$

3. Conditions for overlap:
   $cOb \wedge cR_+a \wedge bR_-a \rightarrow aOb$

4. Transitivity:
   $aR_+b \wedge bR_+c \rightarrow aR_+c$

5. Conditional transivitiy for O
   $aOc \wedge cOb \wedge cR_+b \wedge cR_+a \rightarrow aOb$

6. Linearity
   $bR_+a \vee aR_+b$

7. Covering axiom
   $\exists c(a \preceq c \wedge b \preceq c)$

8. Substitution principle
   Any sentence obtained from the above axioms by replacing R_- by R_+ and R_+ by R_-.

An *event structure* is a tuple $\mathcal{W} = (W, R_+, R_-, O, \preceq)$ satisfying the above axioms with W finite. A first interesting property of these axioms in the light of the previous chapter is that they are all geometric implications, and only axiom (6) makes use of a disjunction in the consequent. It is also clear that time is 'potentially infinite' in this formalism, since new events can be added between two existing events in a structure *ad infinitum*, resulting in more and more fine-grained temporal models. Importantly, the system can be explained as a "determination of sensibility by the understanding". The first three axioms seem to follow from any intuitive interpretation of 'covering' and 'overlapping', but the correspondence of the latter 4 axioms with the Kantian system needs a bit more elaboration. Transitivity can be defined as a composition of reproduction by the imagination: if a is reproduced in the present during which b is produced (aR_-b) and b is reproduced in the present during which c is produced (bR_-c), then these reproductions must necessarily be composed so that aR_-c. This necessity of composition is grounded by the category of *causality*, of which Kant states that it determines succession in time, and which itself conveys transitivity in chains of causation. The linearity axiom insists that any two events are comparable as existing in the same time. No two temporal acts of synthesis could produce distinct timelines. This axiom follows from the category of *community*, under which "all appearances, as contained in a possible experience, must stand" (A214/B261). This category is often considered as ground for simultaneity through reciprocal influence, but can in the same manner determine the community of all appearances in one-dimensional time (A189/B232). The covering axiom is unique in positing the existence of additional events from *a priori* grounds. It is grounded in the category of substance as "that which persists, ..., the substratum of the empirical representation of time itself" (A183/B226). This axiom turns the preorder $\preceq$ into a directed preorder, which for finite structures is equivalent to the existence of a maximal element or *universal cover*, i.e., a representation of unbounded time. The substitution axiom expresses a symmetry between past and future, so that any *event structure* might also be interpreted as a 'reversed' event structure where past and future are switched. The choice for a specific orientation is however determined by causal laws, as is clear from the example behavior provided at the end of this chapter.

The construction provided by Pinosio further contains the binary operators $\oplus$ and $\ominus$ which produce causal futures and causal pasts. Intuitively, $a \oplus b$ is the unique maximal event that begins after both a and b, and ends simultaneously with a, if such an event exists. It thus represents the part of a that can be causally influenced by b. Since the introduction of these operators results in a combinatorial explosion of events, they are not included in the FAE. Instead, causal influence is expressed in terms of the temporal relation between events directly.

The objective for the FAE is then to represent sensation in time by constructing *event structures*. It will do so starting from either a sequence of atoms associated with natural numbers, or a *partial event structure*, where the latter is a tuple $\mathcal{W} = (W, R_+, R_-, O, \preceq)$ which does not need to satisfy any of the temporal axioms. The missing atomic relations may then be decided either positively or negatively to produce a full event structure. One may easily see that there always exists an event structure $\mathcal{W}' = (W', R'_+, R'_-, O', \preceq')$ such that the elements of $\mathcal{W}$ are subsets of the elements of $\mathcal{W}'$, since none of the axioms enforce the negation of an atomic formula. I then call the latter an extension of $\mathcal{W}$ and write $\mathcal{W} \subseteq \mathcal{W}'$. Whether there exists a *sensible* extending event structure, i.e. one where lamps are not on and off at the same time, is however not at all certain, and must be determined by the FAE. Before going through examples of input and sensible temporal structures, I must however first say more about the structure of time and unification from a mathematical point of view.

Boundaries and intervals

An additional concept featuring in the temporal system introduced above is that of a *boundary* (Pinosio, 2017, p.115). Boundaries are an important tool to represent moments as separating connected limitations of time.

Definition 10. *A **boundary** is a tuple (P, C, F) of sets of events, where:*

1. $P, C, F \subseteq W$
2. $F \cup C \cup F = W$
3. *If* $P = \emptyset$ *or* $F = \emptyset$ *then* $C = \emptyset$
4. P *is closed under* R_-
5. F *is closed under* R_+
6. *No events in* P *and* F *overlap*
7. *Every event in* C *overlaps with an event in* P *and an event in* F.

Intuitively, we can think of boundaries as dividing event structures in a past, current and future. A boundary thus delimits past and future, but condition 7 ensures that it also links the two. Furthermore, for every pair of temporally separated events, i.e., non-overlapping events a, b such that aR_-b, there exists a boundary (P, C, F) with $a \in P, b \in F$ (Pinosio, 2017, p.114). If the latter holds, we also say that a and b are separated by a boundary. We can then interpret boundaries as the moments that Kant described as "succession of impressions", following his claim that time consists of times and "points and instants are only boundaries, i.e., mere places of their limitation" (A170/B212). Now given an *event structure* $\mathcal{W}$, we can consider the set of all its boundaries $\mathcal{B}(\mathcal{W})$ and order them so that $(P, C, F) \leq (P', C', F')$ iff $P \subseteq P'$. Pinosio has then shown that the $\leq$ relation is a total linear order, i.e. a total relation that is reflexive, transitive and antisymmetric (Pinosio, 2017, p.116).

Given the representation of moments or 'places of limitation' as boundaries, we may investigate the times that they delimit. Pinosio denotes these as *intervals*. Boundaries and intervals are jointly referred to as *instants*:

Definition 11. *An **instant** is a tuple (P, C, F) of events where:*

1. *$P, C, F \subseteq W$*

2. *$P \cup C \cup F = W$*

3. *P is closed under R_-*

4. *F is closed under R_+*

5. *No events in P and F overlap*

6. *Every two events in C overlap*

7. *$P \cap C = \emptyset$ and $F \cap C = \emptyset$*

8. *For any $a \in C$ there is b such that aOb and $\neg bOc$ for all $c \in P$*

9. *For any $a \in C$ there is b such that aOb and $\neg bOc$ for all $c \in F$*

One can then check that all boundaries are instants. Furthermore, it can be shown that an instant is not a boundary if and only if $F \subset \{w \in W : \neg\exists p \in P\ wOp\}$ (Pinosio, 2017, p.120). If the latter condition holds, we instead name our instant an *interval*. We can define a jump between boundaries as a pair of boundaries $(x, y) \in \mathcal{B}(\mathcal{W})$ such that $x < y$ and there exists no z such that $x < z < y$. Then importantly, the jumps between boundaries are in one-to-one correspondence with the intervals:

Proposition 1. *Given event structure $\mathcal{W}$ with boundaries $\mathcal{B}(\mathcal{W})$, there exists a bijection j from the jumps to the set of intervals such that $j(x, y) = (P_x, C, F_y)$, where $C = (P_x \cup F_y)^c$.*

Proof. In (Pinosio, 2017, p.122) □

Hence, every interval is the fleeting time between two adjacent points. As is the case for boundaries, there exists a complete linear ordering between the intervals $\mathcal{I}(\mathcal{W})$ where $i \leq i'$ iff $P_i \subseteq P_{i'}$. We can also find a complete linearly ordered lattice of the set of instants $\mathcal{K}(\mathcal{W})$ where $x \leq y$ iff $P_x \subseteq P_y \wedge F_y \subseteq F_x$. Under this ordering then, for every two boundaries $x < y$ there exists an interval i such that $x < i < y$, and for every jump (x, y) there is exactly one such i. In this sense the intervals then function as a 'glue' between boundaries, preventing the existence of the 'jumps' or 'clefts' in time that Kant rejected:

"There is nothing simple in appearance, hence no immediate transition from one determinate state (not of its boundary) into another [...] a hiatus, a cleft, is a lack of interconnection among appearances, where their transition is missing." (R.4756, 17:699)

Now returning to our computational context, it is of importance that from an *event structure* $\mathcal{W}$, a numbered sequence of atoms can always be constructed by numbering the intervals in $\mathcal{I}(\mathcal{W})$ according to their linear ordering. Hence, if we allow the FAE to construct sensory sequences and traces as *event structures*, it can make sense of the input in a manner that includes that of the original AE: time as *event structure* can be reduced to time as natural numbers.

4.3 Figurative synthesis and formal intuition

If temporal manifolds are represented as event structures, the *synthesis of apprehension and reproduction* can be represented as a process of constructing, merging, refining and reproducing these event structures. In the previous chapter, geometric logic was introduced as the logic of objectively valid formulas on inverse systems, where the inverse limit constitutes the transcendental object. This framework can also be applied to event structures. Pinosio constructs an inverse system of event structures using *retractions*: functions $f : \mathcal{W} \to \mathcal{W}'$ from event structures to substructures such that $f(a) = a$ and such that all relations $R_+, R_-, O, \oplus, \ominus$ are preserved. Proving that any two event structures $\mathcal{W}, \mathcal{W}'$ are subsumed by a third event structure $\mathcal{W}''$ from which there exist retractions to $\mathcal{W}$ and $\mathcal{W}'$, he shows that any collection of event structures can be unified into a single temporal structure $\mathcal{W}_U$. The inverse limit of such a system can then be associated with the *formal intuition* of time as transcendental object. Since the axioms given in the previous section are structured in geometric logic this then gives a very satisfactory representation of objective temporal structure through figurative synthesis. This approach hinges however on the result that retractions always exist from event structures to substructures. Since I do not make use of the operators $\oplus$ and $\ominus$, this is unfortunately no longer the case. This is shown by figure 4.1:

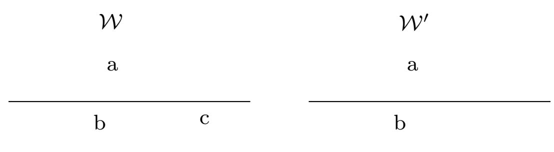


Figure 4.1: $\mathcal{W}'$ is a substructure of $\mathcal{W}$, but there exists no retraction from $\mathcal{W}$ to $\mathcal{W}'$.

Since aR_-c and cR_+b, there is no event in $\mathcal{W}'$ to which c can be mapped while preserving the relational structure. I thus opt for an alternative approach. Instead of constructing an inverse system I represent synthesis of apprehension and reproduction for the FAE as a system of embeddings and associate time as *formal intuition* with the *direct limit* of this system. Embeddings are then functions $f : \mathcal{W}' \to \mathcal{W}$ such that for every relation $R \in \{R_+, R_-, O, \oplus, \ominus\}$: $\mathcal{W}' \models R(a_1, a_2)$ iff $\mathcal{W} \models R(f(a_1), f(a_2))$. The 'Homomorphism Preservation Theorem' then applies (Rossman, 2008) :

Theorem 2. *A first order formula is preserved under finite homomorphisms on all structures if and only if it is equivalent to a geometric formula.*

The truth of geometric formulas is thus preserved in the process of embedding event structures. Now to guarantee a form of global consistency, the embeddings must satisfy a condition of amalgamation:

Definition 12. *A system of event structures related by embeddings satisfies the* amalgamation condition *if for any event structure $\mathcal{W}$ with embeddings $f : \mathcal{W} \to \mathcal{V}$ and $g : \mathcal{W} \to \mathcal{U}$ there exist event structure $\mathcal{K}$ and embeddings $f' : \mathcal{V} \to \mathcal{K}$, $g' : \mathcal{U} \to \mathcal{K}$ so that $f'(f(a)) = g'(g(a))$ for all $a \in W$.*

Let $\mathcal{W} \models \Box\phi(d)$ if for all event structures $\mathcal{W}'$ such that embedding $f : \mathcal{W} \rightarrow \mathcal{W}'$ exists it is the case that $\mathcal{W}' \models \phi(f(d))$, and dually define $\Diamond$. Then note that if we assume the existence of reflexive embeddings, by the preservation theorem $\phi \leftrightarrow \Box\phi$ holds for geometric formulas. Amalgamation now ensures that $\Diamond\Box\phi \rightarrow \Box\Diamond\phi$ holds for each event structure and formula ϕ. I further insist that there exists a maximal element of the structure, which together with the *amalgamation condition* ensures unification:

Definition 13. *A set of event structures related by embeddings satisfies the* unification condition *if and only if for each event structure and formula ϕ: $\Box\Diamond\phi \rightarrow \Diamond\Box\phi$.*

Note that I do not require ϕ to be geometric, since for geometric ϕ the formula $\Box\Diamond\phi \rightarrow \Diamond\Box\phi$ directly follows from reflexivity and $\phi \rightarrow \Box\phi$, without the need for amalgamation and a maximal element. In what follows the maximal temporal structure is termed $\mathcal{W}_U$; the result of figurative synthesis, which is to be unified by conceptual judgement. It is again a finite event structure, since it is constructed from merging finitely many finite event structures and only adding finitely many additional events at each embedding.

Unification through embeddings for the FAE

I now more concretely address the system of embeddings constructed by the FAE, and subsequently the construction of the *direct limit*, as alternative for the *inverse limit*. The FAE starts with a set of 'initial' partial event structures. The event structures are partial because only some of the atomic relations in these event structures have been determined by the input, and the complete event structured are produced later in the process of temporal unification. The term 'initial' here means that the event structures are the starting point of the system of embeddings: each is embedded in another event structure, but none of them is itself the image of a (non-reflexive) embedding. From these initial partial event structures a system of embedded event structures is constructed, with $\mathcal{W}_U$ as maximal element. Throughout this system, each non-initial event structure $\mathcal{W}$ is the image of two embeddings $f' : \mathcal{W}' \rightarrow \mathcal{W}, f'' : \mathcal{W}'' \rightarrow \mathcal{W}$, where $\mathcal{W}', \mathcal{W}''$ are substructures of $\mathcal{W}$. Furthermore, if input is given with a sequence of natural numbers, $\mathcal{W}'$ and $\mathcal{W}''$ are disjoint except for their cover event, which must exist due to the 8th temporal axiom. This cover represents the whole time-span of synthesis and is shared throughout all event structures in the system. If input is given as partial event-structure, $\mathcal{W}'$ and $\mathcal{W}'''$ might however share more events. The larger system $\mathcal{W}$ must contains all events in $\mathcal{W}'$ and $\mathcal{W}''$, but it typically also contains additional 'merged' events that are added to unify the content of $\mathcal{W}'$ and $\mathcal{W}''$ in a manner that allows for conceptual interpretation. All embeddings are identity functions, i.e. $f'(e) = e$ for all $e \in W'$, which together with the property that all successors share a common successor and the existence of a maximal element ensures satisfaction of the *amalgamation* and *unification* conditions. The precise mechanism that constructs embeddings will be made more clear in section 4.5. For now it is important however that we should not understand the FAE as constructing embeddings bottom-up from existing event structures as starting point. Instead, the whole system of event structures is constructed in parallel. The search for a conceptual theory has a top-down influence on the structure of $\mathcal{W}_U$ and hence on that of all event structures in the system, from which there must exist embeddings to $\mathcal{W}_U$. Conversely, the partial structure of event structures that follows from the input has a bottom-up effect on $\mathcal{W}_U$ and hence on the structure of the conceptual theory that makes sense of the input.

The reader might note that such a parallel construction of embeddings is opposed to the successive nature of synthesis emphasized by Kant (A189). For example, one might construct an event structure where an event a covers the whole duration of the synthesized time, such as $\mathcal{W}_{U1}$ in figure 4.2. Now if one successively senses an event with a content that cannot occur simultaneously with a, say $\neg a$, a problem occurs. Every embedding of $\mathcal{W}_{U1}$ and $\mathcal{W}_2$ into a merged event structure $\mathcal{W}_{U2}$ such that t_1 is mapped onto the cover of $\mathcal{W}_{U2}$ must also have a as cover, contradicting that a does not overlap with $\neg a$. In words: if one wants to add new senstations to the existing temporal structure by means of embeddings, one might have to re-evaluate what was synthesized before.

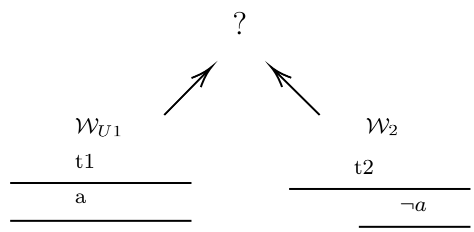


Figure 4.2: Issue with successive bottom-up embedding

In response to this issue, we must firstly be aware that the computational system currently under consideration is still far off from the successive synthesis that Kant had in mind. As noted before, the ASP framework requires us to provide a sensory sequence as a single set. If we want to add an additional sensory atom, we have to begin the synthesis wholly anew. Consequently, in the process of synthesis the FAE has the freedom to restrucure all event structures so as to fit in $\mathcal{W}_U$, and it does not need to take some event structures as 'given' starting points for a bottom-up construction. This then explains why this theoretical tension with Kant's conception of synthesis is not an issue for the functionality of the FAE. Secondly, we can evaluate the mathematical framework by itself: does this issue imply that the proposed system of embeddings can never represent Kant's successive synthesis? In this context we may note that the only event that prevents the embeddings from simply performing a union operation is the temporal cover, that is shared across all event structures. Hence, if one aims to embed new sensations in a temporal structure, while *reproducing* what was synthesized before, the only necessary adjustment would concern the representation of time as a whole. If the synthesis covers a larger time-span, the relation between time and the manifold of sensation has to be adjusted accordingly, but the temporal orientation of events corresponding to the reproduced sensations may remain the same. The adjustment of structure that may be necessary in the process of reproduction thus seems to me to be both sensible and manageable.

Direct limit and formal intuition

Now finally, I show how this system of embeddings can be intepreted as a partially ordered set with a direct limit, following the general definition of direct limits on partially ordered sets of structures (Hodges, 1993, Ch.2). Since the composition of two embeddings is an embedding, and since all embeddings are identity functions, we can define an embedding unambiguously between every two event structures $\mathcal{W}, \mathcal{W}'$ between which there exists a sequence of embeddings as the composition of the sequence. If we then additionally add the identity relations that embed each event structure in itself, the result is a partially ordered set of event structures $(E, \leq)$ where $\mathcal{W} \leq \mathcal{W}'$ denotes the existence of an embedding $f_{\mathcal{W}\mathcal{W}'} : \mathcal{W} \to \mathcal{W}'$ in the system. Now following the structure of direct limits presented by Hodges, we can write X for the union over the domain of all event structures, which is equal to the domain of $\mathcal{W}_U$. Then, we define an equivalence relation $\sim$ as follows:

$$\text{If } a \in dom(\mathcal{W}), b \in dom(\mathcal{W}'), \text{ then } a \sim b \text{ iff there exists } \mathcal{W}'' \text{ such that } \mathcal{W} \leq \mathcal{W}'', \mathcal{W}' \leq \mathcal{W}'' \text{ and } f_{\mathcal{W}\mathcal{W}''}(a) = f_{\mathcal{W}'\mathcal{W}''}(b)$$

In our context, this means $a \sim b$ holds if and only if $a = b$ so that $a^{\sim} = \{a\}$, since all embeddings are identity functions. Now the direct limit B is defined over the set of equivalence classes of $\sim$. Then for each system in the structure of embeddings $\mathcal{W}$ we define an embedding $h_{\mathcal{W}} : W \to dom(B)$ by letting $h_{\mathcal{W}}(e) = e^{\sim}$. Now to ensure that the h maps are embeddings we require:

$$\bar{b} \in R^B \text{ iff there exists } \mathcal{W} \in E \text{ and } \bar{a} \in W \text{ such that } h_{\mathcal{W}}(\bar{a}) = \bar{b} \text{ and } \mathcal{W} \models R(\bar{a})$$

This then in general gives the direct limit of a system of partially ordered structures, a consistent representation of the structure of temporal structures as a whole, i.e., time as *formal intuition*. Since, in the case of the FAE, each event structure in the system is a substructure of $\mathcal{W}_U$, and since we have $a^{\sim} = \{a\}$ the equation above amounts to $\bar{b} \in R^B$ iff there exists $\bar{a} \in W_U$ such that $h_{\mathcal{W}_U}(\bar{a}) = \bar{b}$ and $\mathcal{W}_U \models R(\bar{a})$ and $\bar{b}$ consists of the singleton sets of elements of $\bar{a}$. We can thus simply copy $\mathcal{W}_U$ and replace all elements in the domain of W_U by their singleton sets to achieve the desired limit B: the unified event structure representing *formal intuition*. Going back to what was said on *form of intuition* and *formal intuition* in the introduction, we then have the *form of intuition* as the temporal axioms that convey *unactualized* temporal structure with which all synthesis in time must be in agreement. The system of reflexive embeddings under conditions of *amalgamation* and *unification* can on the other hand be considered the source of combination that applies to the manifolds of intuition. From its pure structure this system ensures the existence of *formal intuition* as the unified whole $\mathcal{W}_U$, but in its empirical application the same structure represents the manifold of intuition *as manifold* in time as do the *syntheses of apprehension and reproduction*. Note that $\mathcal{W}_U$ is a single structure that can be defined by a single conjunction of atomic facts, highlighting the objective nature of *formal intuition* as unity through necessary connection. We will further see that $\mathcal{W}_U$ is pre-conceptual in the sense that it is only the starting point of conceptual unification, but is also determined by this conceptual understanding, so that we may rightly name it the result of a pre-conceptual synthesis.

4.4 Topological interpretation

Before turning to the computational implementation, I discuss a topological interpretation of event structures that sheds light on their correspondence with the Kantian system. First, we need some definitions. As usual I denote with a topology τ on set X a family of subsets of X (i.e., $\tau \subseteq \mathcal{P}(X)$) that is closed under finite intersections and arbitrary unions such that $\emptyset \in \tau, X \in \tau$. Sets in the topology are open, and complements of open sets are closed. An Alexandroff topology is a topology that is additionally closed under infinite intersections. Now let X be a preorder, i.e., the elements of X are related by a relation that is reflexive and transitive, then the corresponding Alexandroff topology $\mathcal{A}_{\leq} \subseteq \mathcal{P}(X)$ is defined as follows:

Definition 14. *The Alexandroff topology $\mathcal{A}_{\leq} \subseteq \mathcal{P}(X)$ is the set of all upsets of X, where an upset is a set $\{x \in X : x \geq s \text{ for some } s \in S\}$ generated by a set $S \subseteq X$.*

Conversely, given an Alexandroff topology $\tau \subseteq \mathcal{P}(X)$. One can define the specialization ordering $\sqsubseteq^{\tau} \subseteq X \times X$ by letting $x \sqsubseteq y$ if $U \in \tau \wedge x \in U$ implies $y \in U$. Then the following theorem was proven by Alexandroff (1937):

Theorem 3. *For all preorders $\leq$ on a set X, $\sqsubseteq^{\mathcal{A}_{\leq}} = \leq$, and for all Alexandroff topologies τ: $\mathcal{A}_{\sqsubseteq^{\tau}} = \tau$*

Now note that the relations $R_+, R_-, \preceq$ are preorders: reflexive and transitive relations, by axioms (1) and (4). Hence, given an event structure, we can define Alexandroff topologies $\mathcal{A}_+ = \mathcal{A}_{R_+}, \mathcal{A}_- = \mathcal{A}_{R_-}, \mathcal{A}_{\preceq}$, where the open sets are the sets closed under the relevant relation. Note that in the context of the previous section, the sets P, C, F of instants are open sets of $\mathcal{A}_-, \mathcal{A}_{\preceq}$ and $\mathcal{A}_+$ respectively.

We can make use of this topological interpretation to analyse our temporal structure with respect to Kant's claim that time is connected, i.e. parts of time, which are themselves times (A32), "can be distinguished, but not separated" (R.4425, 17:541) and "every determinate magnitude of time is only possible through limitations of a single time grounding it"(B48). In topology, connectedness is usually defined as the condition that no two disjoint non-empty open sets (or equivalently two disjoint closed sets) together form the whole space. Now following Pinosio (2017), we may compare several topological connectedness conditions on event structures with Kant's notion of non-separable time. A first observation is then the following:

Lemma 4. *For any two open sets A, B of $\mathcal{A}_+$, either $A \subseteq B$ or $B \subseteq A$, and the same holds for open sets of $\mathcal{A}_-$.*

This directly implies that the spaces $\mathcal{A}_+$ and $\mathcal{A}_-$ are connected, but these are only the distinct spaces of 'future' sets and 'past' sets. A more interesting result compares both topologies:

Definition 15. *Event structure $\mathcal{W}$ is* biconnected *if there are no non-empty sets $U \in \mathcal{A}_+, V \in \mathcal{A}_-$ such that $U \cap V = \emptyset$ and $U \cup V = W$*

Lemma 5. *All finite event structures $\mathcal{W}$ are biconnected*

Proof. In (Pinosio, 2017), for event-structures with $\oplus, \ominus$, but easily seen to hold here as well. By the covering axiom, there exists a maximal event t such that $e \preceq t$ for all $e \in W$ (a universal cover for $\mathcal{W}$). We have $t \in U$ or $t \in V$, implying $U = W$ or $V = W$ so that $V = \emptyset$ or $U = \emptyset$, a contradiction. □

An even stronger notion of connectedness however makes use of the join topology $\mathcal{A}_+ \vee \mathcal{A}_-$, which is the topology that has $\mathcal{A}_+ \cup \mathcal{A}_-$ as subbase (i.e., $\mathcal{A}_+ \vee \mathcal{A}_-$ is the smallest topology containing $\mathcal{A}_+ \cup \mathcal{A}_-$). The open sets of $\mathcal{A}_+ \vee \mathcal{A}_-$ are the closed sets of $\mathcal{A}_\preceq$. The closed sets are conversely the upsets of $\preceq$. The existence of a universal cover then also directly implies the following, since any two closed sets of the join topology contain the representation of time as a whole:

Definition 16. *A topological space is* ultra-connected *if any two non-empty closed sets have non-empty intersection.*

Lemma 6. *Event structures with the $\mathcal{A}_+ \vee \mathcal{A}_-$ topology are ultra-connected.*

This provides a formal correlate for Kant's claim that time cannot be made up of its parts. Given event structure $\mathcal{W}$ we can identify the $\preceq$ downset generated from event $a \in W$ ($\downarrow_\preceq a$) as a part of $\mathcal{W}$. Note that this is an open set of the join topology, and is also itself an event structure, representing Kant's claim that parts of time are themselves times. The following proposition then follows from the ultra-connectedness of $\mathcal{A}_+ \vee \mathcal{A}_-$:

Proposition 2. *Let $\mathcal{W}$ be an event structure with the $\tau = \mathcal{A}_+ \vee \mathcal{A}_-$ topology. Consider open sets $U, V \in \tau$ such that $W = U \cup V$. Then either $U = W$ or $V = W$. Thus, time is not a union of its proper parts.*

Proof. Suppose $U \neq W, V \neq W$. Then the closed sets $W \setminus U, W \setminus V$ are non-empty, so that they must by ultra-connectedness have non-empty intersection, contradicting that $U \cup V = W$. □

Note that ultra-connectedness hinges on the universal cover. This warrants an interpretation of this universal cover as "representation of time itself" (A183/B226), of which the parts are only given by limitation (B48). It is then by virtue of time as a whole, as represented within the *formal intuition* $\mathcal{W}_U$, that the structure of time is a unity that cannot be made up of its parts. The covering axiom and the structure of embeddings introduced in section 4.3 ensure that this universal cover is indeed constructed as a pure object: the result of synthesis *a priori*. We have now associated two mathematical entities with Kant's *times in time*. Intervals are minimal extended entities delimited by two boundaries. These fleeting times between moments ensure the flowing of time and prevent the existence of clefts. The downsets of $\mathcal{A}_\preceq$ are instead parts of time with the same mathematical structure as time as a whole (i.e. an event structure). They show how a temporal extension may be divided into smaller extensions, but only so that the parts cannot make up the whole. Each of the two mathematical notions thus highlights a distinct aspect of Kant's views on temporal connectedness.

4.5 Implementation

The temporal structure introduced above is fundamentally different from that of the original AE. The implementation thus takes the form of a wholly new system, rather than an extension. I hence provide a more coarse-grained overview than in the previous chapter, leaving out the precise coding structure in some cases and in other cases providing code blocks without an extensive explanation. For the reader experienced with ASP some of the programs might seem unnecessarily complicated. In these cases, the coding structure was almost always prompted by reasons of computational efficiency: I have restricted the domain of choice where possible to make the burden for the system lighter, thereby making the burden for the reader heavier. Firstly, I explain how the unified event structure $\mathcal{W}_U$ is constructed from the input. Then, I turn to the notion of 'making sense', which now corresponds to the covering of $\mathcal{W}_U$ by a conceptual event structure $\mathcal{W}_T$ generated by theory T.

4.5.1 Events and Manifolds

Starting with input as a sequence of sensory atoms with natural numbers, each atom is associated with an 'event' representing an act of apprehension and reproduction. Events occurring at the same position in the input ordering are placed in a 'manifold', which is an event structure. The system thus constructs a 'manifold of manifolds' which "distinguishes the time in the sequence of one impression upon another" (A99). The event structures constructed in this manner are then embedded or 'merged' into larger event structures as explained in section 4.3. Events are numbered and manifolds are represented as tuples (x, y), where x represents the input ordering index and y represents the depth in the process of combination. The events constructed from the input are thus initially placed in manifolds $(t, 0)$ where t represents the input index:

```
input_size(X) :- X = #count {C ,T : senses(C, T)}.

max_time(X) :- X =  #max {T : senses(_, T) }.

subjective_time(1).
subjective_time(I+1) :- subjective_time(I), max_time(X), I<X.

pos_intuit_event(1..2*X) :- input_size(X).

sensed_atoms_at_time(T,C) :- subjective_time(T),
    C = #count {X : senses(X,T)}.

sensed_upto(1,0).

sensed_upto(T+1, C+X) :- sensed_upto(T, C),
    sensed_atoms_at_time(T,X), subjective_time(T+1).

% Assign 'intuition' to events with the same content,
% using 'sensed_upto' to restrict the domain of choice.
sense_input_event(E, Content, T) : pos_intuit_event(E), E > X, E <= X+Y :-
    senses(Content, T), sensed_upto(T, X),
```

```
    sensed_atoms_at_time(T, Y).

% Assign events to manifolds.
has_event((T,0), E) :- sense_input_event(E, _, T).

intuit_event(E) :- has_event(M, E),
    pos_intuit_event(E).
```

Program 4.1: Event construction and assignment to manifolds

Here, 'pos_intuit_event' identifies the domain of events available to the system, which is set at twice the size of the input, leaving room for the process of combination in which additional events can be constructed.

4.5.2 Time for events

To place the events in a temporal structure, a sequence of choice rules and constraints implements the axioms introduced in section 4.2. The relations R_+, R_- and O are named 'r_after', 'r_before' and 'time_overlap' and are generated by means of choice rules:

```
1 {r_after(E1, E2); r_after(E2, E1) } :- event(E1), event(E2).

1 {r_before(E1, E2); r_before(E2, E1) } :- event(E1), event(E2).

{time_overlap(E1,E2)} :- event(E1), event(E2).
```

Program 4.2: Generation of temporal relations

The first two rules also implement the linearity axiom (6). Axioms (1) up to (5) and their counterparts produced by substitution are straightforwardly represented by constraints on the relations, which is not shown here. The covering axiom (7) is implemented by construction of a universal cover, the object representing time as a whole that covers all intuition in time:

```
time_overlap(E1, E2) :-
    sense_input_event(E1, _, T), sense_input_event(E2, _, T).

time_cover_event(2*X+1) :- input_size(X).

time_covers(T,E) :- time_cover_event(T), intuit_event(E).
```

Program 4.3: Construction of the temporal cover

I further define relations 'r_strictly_before' as $R_-(a,b) \wedge \neg R_-(b,a)$ and 'r_strictly_after' likewise.

4.5.3 Contents

Events are given a content as well as a time. This results in an important distinction with the system used by Pinosio, in which events with the same time are made identical by an 'extensionality' axiom:

$$a \preceq b \wedge b \preceq a \rightarrow a \equiv b$$

I instead allow events at the same time to be distinguished on the basis of their contents, and insist that events can have at most one content. Since events are extended in time, their content is however not a single formula in predicate logic, but an 'alteration' in the sense of Kant's first analogy of experience. For the FAE, alterations are either changes in properties represented by a unary predicate (for instance, $on(light)$ changes to $off(light)$, represented by $(light, (on, off))$), changes represented by a binary predicate (for instance $close(agent, wall)$ changes to $far(agent, wall)$, represented by $(agent, wall, (close, far))$), or changes in a numerical value associated with a binary predicate (for instance $temp(water, 2)$ changes to $temp(water, 5)$, represented by $(temp, water, (2, 5))$). Each content is characterized by the states at the beginning and end of the event, or its 'beginning content' and 'ending content'. If the beginning and ending content are the same an event is called *stable*. The events constructed directly from the atoms in the input are then always stable, since alterations are only constructed in the process of combination:

```
has_content(E, s(O,(C,C))) :-
    sense_input_event(E, s(C,O), _).

has_content(E, s2(O1,O2,(C,C))) :-
    sense_input_event(E, s2(O1,O2,C), _).

has_content(E, s2v(C,O,(V,V))) :-
    sense_input_event(E, s2v(C,O,V), _).
```

Program 4.4: Derivation of contents from input

I provide a number of axioms to enforce consistency on events and their contents. For instance, 'incompossible' events that exclude each other cannot overlap in time, events that start at the same time and refer to the same object and property must have the same beginning content (where for instance 'on' and 'off' are different values for the same property), and different directions in numerical alterations cannot overlap (e.g., temperature cannot rise and fall at the same time). Since this requires quite a long program, a complete overview of these constraints is given in the appendix:

4.5.4 Input as partial event structure

I now show how the FAE can take partial event structures as input, so that no counting of times has to be done *before* the reception of sensation in sensibility. As explained before, partial event structures do not yet need to satisfy all temporal axioms. I thus simply provide a set of atomic relations between sensations as input such as:

$$
\begin{aligned}
&R_+ \ ((light, on), 1), ((light, off), 1) \\
&O \ (light, on), 1), ((light, off), 1) \\
&R_- \ ((above, lamp, table), 1), ((light, off), 2) \\
&\ldots
\end{aligned}
$$

The numbers 1, 2 are here merely used to distinguish distinct tokens of the same type, i.e. distinct sensations with the same content. We might just as well have used a, b, c or $-100, 3.4$ etc.. Given this partial event structure, the FAE must distinguish the sensory atoms and produce a complete event structure to construct its own temporal succession:

```
input((C,I)) :- r_before((C,I),_).
input((C,I)) :- r_before(_,(C,I)).
input((C,I)) :- r_after((C,I),_).
input((C,I)) :- r_after(_,(C,I)).
input((C,I)) :- time_overlap((C,I),_).
input((C,I)) :- time_overlap(_,(C,I)).

1 {r_before(I1, I2); r_before(I2, I1) } :-
    input(I1), input(I2).

1 {r_after(I1, I2); r_after(I2, I1) } :-
    input(I1), input(I2).

{time_overlap(I1,I2)} :- input(I1), input(I2).
```

Program 4.5: Construct an event structure from the partial structure in the input

Then, the FAE identifies boundaries in this event structure. Since definition 10 is quite cumbersome as a whole I reduce the task to finding pairs of events that are *arbitrarily close* in the following sense:

Proposition 3. *Given event structure $\mathcal{W}$ and boundary $b = (P, C, F)$ such that $P \neq \emptyset, F \neq \emptyset$, there exists at least one event $p \in P$ and $f \in F$ such that $\neg pOf$, pR_-f and no event begins or ends in the space between p and f. I call such p, f* ***arbitrarily close****.*

Proof. We may firstly note that R_+ and R_- form *total preorders* (i.e. relations that are transitive and total) on finite domains. A total preorder $\lesssim$ on a finite domain X always has at least one largest element x such that $\forall y \in X : y \lesssim x$. To see this, we may simply take an arbitrary element $a \in X$. If it is not a largest element, there must by totality exist $b \in X$ such that $a \lesssim b$ and by transitivity $\forall a' \in X$ such that $a' \lesssim a$ we have $a' \lesssim b$. If b is not a largest element we continue this process. Since X is finite the process must halt at a largest element. In our context then, there must exist $p \in P$ such that $\forall x \in P : xR_-p$ and $f \in F$ such that $\forall y \in F : yR_+f$.

Now by properties 4 and 6 of definition 10 we have $\neg pOF$ and $\neg fR_-p$, so that by linearity pR_-f. Now finally, suppose that there exists an event e ending in the space between p and f. Then since $\neg eR_+f$, while f is a largest element of F we have $e \notin F$. Since $\neg eR_-p$, while p is the largest element of P we have $e \notin P$. Thus, $e \in C$. But then by property 7 of definition 10 there exists $f' \in F$ such that eOf'. Since f is a maximal element of F then $f'R_+f$. Furthermore, since e ends before f we have eR_-f. Now we can apply temporal axiom 3 with $a = f, b = e, c = f'$ to derive eOf, contradicting that e ends in the space between p and f. Thus, no event can end in the space between p and f. By a symmetric argument also no event can begin in this space. □

The reader may easily check that furthermore any two arbitrarily close events are separated by a unique boundary. This can be seen directly by defining C as the set of all events overlapping with both p and f. The system thus identifies pairs of arbitrarily close events to find boundaries, and subsequently identifies jumps between pairs of successive boundaries in the ordering of proposition 1

```
exists_end_between(I1,I2) :-
    input(I2), input(I2), input(I3),
    r_after(I2, I1), not time_overlap(I1,I2),
    r_strictly_before(I1,I3), not r_after(I3,I2),
    not time_overlap(I3,I2).

exists_begin_between(I1,I2) :-
    input(I1), input(I2), input(I3),
    r_after(I2, I1), not time_overlap(I1,I2),
    r_strictly_after(I2,I3), not r_before(I3,I1),
    not time_overlap(I3,I1).

arbitrarily_close((I1,I2)) :-
    input(I1), input(I2), r_after(I2, I1),
    not time_overlap(I1,I2), not exists_begin_between(I1,I2),
    not exists_end_between(I1,I2), not bound_after_already_exists(I1),
    not bound_before_already_exists(I2).

bound_after_already_exists(I1) :-
    arbitrarily_close((I3,I4)), I3 != I1,
    r_before(I1,I3), r_before(I3,I1).

bound_before_already_exists(I2) :-
    arbitrarily_close((I3,I4)), I4 != I2,
    r_after(I2,I4), r_after(I4,I2).

less((A1,A2), (B1,B2)) :-
    arbitrarily_close((A1,A2)), arbitrarily_close((B1,B2)),
    r_before(A1,B1), not r_before(B1,A1).

jump(A,B) :-
    less(A,B),
    not less(C,B) : arbitrarily_close(C), less(A,C).
```

Program 4.6: Identify boundaries and jumps

From the jumps between boundaries a linear ordering is constructed and associated with natural numbers. Then finally, given a jump between two pairs of arbitrarily close events $((p_1, f_1), (p_2, f_2))$, any event e such that $\neg eR_{-}p_1 \wedge \neg eR_{+}f_2$ is placed in the associated interval. The number of this interval is then the number of its upper bound. The starting and ending interval are treated appropriately, so that interval 1 lies before the first pair of arbitrarily close events. The intervals now have the same function as the input succession in the previous sections: events are placed in one or multiple manifolds $(x, 0)$ where x is the number of their interval. Importantly, all temporal relations of the input event structure are maintained in the manifolds $(x, 0)$, and thus also in the unified temporal representation $\mathcal{W}_U$. We must note again however that like the construction of embeddings described in section 4.3, this process is *not* properly described bottom-up. The processes of sensibility and understanding are executed jointly, so that the FAE also structures the event structure it builds from the input in a manner that fits with a conceptual theory.

4.5.5 Merging and alteration

Manifolds are combined by means of embeddings as explained in section 4.3. Given two event structures (manifolds) $\mathcal{W}$ and $\mathcal{W}'$, the combined manifold $\mathcal{W}''$ is constructed, where $W \cup W' \subseteq W''$. The sequence of embeddings is such that every two event structures are embedded in a common event structure, and their exists a maximal element, so that the amalgamation and unification conditions from definitions 12 and 13 are satisfied. I have chosen for the structure represented in figure 4.2 to minimize the number of embeddings, but any alternative structure is possible that satisfies the conditions of global consistency. If a succession with natural numbers is given in the input, this constrains the possible orientation of events in the combined manifolds with the following relations:

If e and f are constructed from the same input position,
then eOf.
If e is constructed from an input position after that of f,
then f starts before e ends: $eOf \vee fR_-e$.[1]

Note that if the succession was constructed actively by the FAE as interval these relations already follow directly. Any two events in the same interval overlap (6 of 11). Further, if e is in an interval (P_1, C_1, F_1) after (P_2, C_2, F_2) containing f then either e, f must also be in a shared interval so that eOf, or $f \in P_1$ so that fR_-e. These are then the only logical 'constraints of the search space' associated with the input succession. Note that conversely these logical constraints do not at all determine the temporal succession in the unified representation $\mathcal{W}_U$. Even if an input ordering is given, this is only the starting point of the succession through moments and intervals that the FAE builds. For instance, if e is associated with input position 4 while f is associated with input position 3, it might very well still be the case that e and f cover each other in the final construction, or that e begins before f so that $e \in C_i$ $f \in C_j$ for intervals $i = (P_i, C_i, F_i) < (P_j, C_j, F_j) = j$.

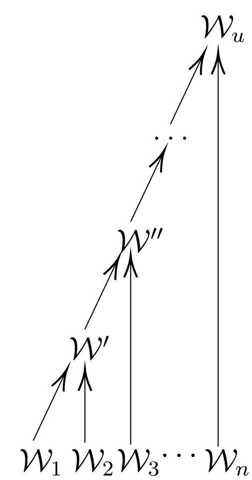


Figure 4.3: System of embeddings constructed by the FAE

In this process of combining manifolds, *alteration events* are constructed as 'merged events'. If two events in the embedded manifolds represent different states of the same object, an event of alteration is assumed to exist between them. Suppose for example that manifold (x_1, y_1) contains event e with content $light(on, on)$ and manifold (x_2, y_2) where $x_1 < x_2$ contains event f with content $light(off, off)$. Note that by the constraints on incompossible contents, e and f cannot overlap. By the above then eR_f since f is after e in the input ordering, i.e. e must end before f starts. There must thus exist an event g with content $light(on, off)$, that starts after e ends, and ends before f begins. Of course, not all events with incompossible contents must be connected in such a way. For example if (x_1, y_1) also has an event e' with content $light(on, on)$ before e, we do not also want to construct an alteration between e' and f. The code that constructs merged manifolds and selects events to merge can be found in the appendix. I only include the program here that constructs contents for merged events and determines the temporal position of events in merged manifolds:

```
time_overlap(E1, E2) :-
    sense_input_event(E1, _, T),
    sense_input_event(E2, _, T), E1 != E2.

% The first argument of 'merged_manifold' and 'merged_event'
% represents the merged object itself, while the latter two
% represent its source objects.
time_overlap(E1, E2) :-
    merged_manifold(A, B, C), has_event(B, E2),
    has_event(C, E2), not r_before(E1, E2).

has_content(E3, s(O,(C2,C3))) :-
    merged_event(E3, E1, E2),
    has_content(E1, s(O,(_,C2))),
    has_content(E2, s(O,(C3,_))).

has_content(E3, s2(O1,O2,(C2,C3))) :-
    merged_event(E3, E1, E2),
    has_content(E1, s2(O1,O2,(_,C2))),
    has_content(E2, s2(O1,O2,(C3,_))).

has_content(E3, s2v(C,O,(V2,V3))) :-
    merged_event(E3, E1, E2),
    has_content(E1, s2v(C, O, (_, V2))),
    has_content(E2, s2v(C,O,(V3,_))).

:- merged_event(E3,E1,_), time_overlap(E3, E1).

:- merged_event(E3,_,E2), time_overlap(E3, E2).

r_strictly_after(E3, E1) :-
    merged_event(E3, E1, _).

r_strictly_before(E3, E2) :-
    merged_event(E3, _, E2).
```

Program 4.7: Assign content to merged events

When this process of merging manifolds has been completed, the sensory sequence has been transformed into a single unified event structure $\mathcal{W}_u$, the result of apprehension and reproduction in time; the representation of the manifold of intuition *as manifold.* Note that if the system is not given an input, $\mathcal{W}_U$ only contains the universal cover: time as a whole. In this sense then, the *formal intuition* of time is the result of *pure figurative synthesis.*

4.5.6 Making sense of event structures: definitions

The next task at hand is now to provide unity under concepts to the pre-conceptual manifold. Judgements must be constructed that explain the sensed input. Since time is however no longer a sequence of natural numbers, we cannot define the extension of 'arrow rules' and 'causal rules' as was done for the original AE. In fact, the whole definition of 'making sense' based on the covering of a sequence $S = (S_1, \ldots, S_n)$ by a trace is no longer applicable. To construct judgements that can explain event structures I define a new language CONTENTLOG that replaces DATALOG. It contains two types of rules: *causal rules* and *regular successions. Causal rules* determine a relation between events of cause (c) and effect (e). In the intended use of CONTENTLOG, the latter is always an *alteration event* which is directly followed by a *stable event* f so that e and f are *arbitrarily close.*

While Kant insists that a causal rule is "this necessitation that first makes possible the representation of a succession in the object" (A197), he also recognises that in many cases cause and effect are simultaneous with one another:

"The majority of efficient causes in nature are simultaneous with their effects, and the temporal sequence of the latter is occasioned only by the fact that the cause cannot achieve its entire effect in one instant. But in the instant in which the effect first arises, it is always simultaneous with the causality of its cause" (A203/B248)

I thus insist that $e \preceq c$: the cause covers the alteration that is its direct effect, and in the instant where e first arises, it is simultaneous with the causality of c. Since the state resulting from the *alteration event* e is represented by the *stable event* f, we might also name f the indirect effect of c. Note than that a causal rule indeed determines a succession between c and f in the sense that f must begin strictly after e begins.

Causal rules in CONTENTLOG are intended to represent efficient causality in nature. They relate causes with their simultaneous effects. The causal rules of the original AE on the other hand determine that a state A is regularly succeeded by state B, but do not identify the cause of the transition between A and B. They identify a 'regular succession', but do not attempt to explain the underlying mechanism of change. This more parsimonious representation of causation is however better suited to the tasks solved by the original AE, which are generally of a logical or game-theoretic nature. The behavior of cellular automata for instance, follows a system of regular succession between states, without any underlying mechanism of causation. To allow the FAE to still solve problems of this kind, I also include rules of 'regular succession'.

Definition 17. *An* unground content *is of the form* $(x, (p, q))$, $(x, y, (p, q))$. $(p, x, (v_1, v_2))$, (p, x, up), *or* (p, x, down) *where* x, y *are variables,* p, q *are predicate symbols,* v_1, v_2 *are constants representing numbers and 'up' and 'down' are constants with a fixed interpretation.*

Definition 18. *A rule in* CONTENTLOG *is of one of two forms:*

$$\alpha_0 \leftarrow_{RS} \alpha_1$$
$$\alpha_0 <<_{CR} \alpha_1$$

Where α_0, α_1 *are* unground contents. *A rule of the first form is a* regular succession*, and a rule of the second form is a* causal rule.

Programs are then sets of rules as usual. A strong limitation is of course that these rules only allow for a single body atom. This significantly eases the application of the language to event structures since events have a single content, but might very well be resolved in later work. I now provide alternative definitions for theories, traces and 'making sense':

Definition 19. *A theory for the* FAE *is a three-tuple* $T = (\phi, I, R)$ *where:*

1. ϕ *is a type signature* (T, O, P, V) *as for the original* AE.
2. I *contains ground atoms, representing the partial state of the world at the beginning of time.*
3. R *is a set of rules in* CONTENTLOG*, where the heads of causal rules always represent alterations, the heads of regular successions only have contents of the form* $(x, (p, q))$, $(x, y, (p, q))$ *where* x, y *are variables, and no two causal rules have the same bodies.*

The last few conditions in (3) are added to ensure that theories provide satisfying explanations, although one might argue for different conditions in different applications. I insist that changes in numerical values can never be explained by the *regular successions* that were included to solve problems of a logical nature, so that the FAE must provide a causal mechanism to explain them. I further insist that no two causal rules may have the same body to persuade the FAE to develop fine-grained explanations. For instance, if a boat passes under a bridge, this causes the bridge to open, which causes traffic lights to switch to red, which causes cars to stop etc.. If I allow the same body to occur in multiple causal rules the FAE can explain all these effects from the single cause 'a boat passes under a bridge', without identifying the more fine-grained mechanism.

Definition 20. *Given a theory* $T = (\phi, I, R)$*, its trace is an event structure* $\mathcal{W}_T$ *where:*

1. *For each atom* α *in* I*, there exists an* init event $e \in W_T$ *with stable content corresponding to* α*, such that* $\forall f \in W_T : f R_+ e$.
2. *For each ground instance* r *of a regular succession in* R*, if there exists an event* $e \in W_T$ *with the body of* r *as content, then there exists an alteration event* $f \in W_T$ *with the head of* r *as content, unless the head of* r *is of the form* (p, o, up) *or* (p, o, down)*, in which case* f *has content* $(p, o, (v_1, v_2))$ *so that* $v_1 < v_2$ *or* $v_2 > v_2$ *respectively. Further,* $f R_+ e$ *and* e *and* f *are arbitrarily close.*

3. *For each ground instance r of a causal rule in R, if there exists an event $e \in W_T$ with the body of r as content, then there exists an event $f \in W_T$ with the head of r as content, unless the head of r is of the form (p, o, up) or (p, o, down), in which case f has content $(p, o, (v_1, v_2))$ so that $v_1 < v_2$ or $v_2 > v_2$ respectively. Further, $f \preceq e$.*

4. *For each* alteration event $e \in W_T$, *there exists a* stable event $f \in W_T$ *such that the ending content of e is the beginning content of f, fR_+e, and e and f are arbitrarily close.*

5. *W_T only contains events of which the existence is warranted by conditions (1) to (4), under the constraint that every event has a single content.*

6. *For every event $e \in W_T$ that is not an* init event, *there exists a unique predecessor event $f \in W_T$ such that the end of the content of f is the beginning of the content of e, eR_+f and f and e are arbitrarily close.*

7. *If event $e \in W_T$ with content $c_e = (o, (c1, c2))$ overlaps with event $f \in W_T$ with content $(o, (c3, c4))$ such that $c3, c4$ represent the same property as $c1, c2$, then eR_+f implies $c1 = c3 \vee c1 = c4$ and eR_-f implies $c2 = c3 \vee c2 = c4$.*

8. *If event $e \in W_T$ with content $c_e = (o1, o2, (c1, c2))$ overlaps with event $f \in W_T$ with content $(o1, o2, (c3, c4))$ such that $o1, o2$ are objects and $c3, c4$ represent the same property as $c1, c2$, then eR_+f implies $c1 = c3 \vee c1 = c4$ and eR_-f implies $c2 = c3 \vee c2 = c4$.*

9. *If event $e \in W_T$ causes event $f \in W_t$, then e does not cover any predecessor of f in the sense of condition (6).*

Condition (6) is the analogue of the frame axiom in the original AE. It ensures that an event e with content $(color, (green, green))$ continues until an event f with content $(color, (green, x)$ starts. Here x is then a concept that is incompatible with 'green'. In this manner the frame axiom and constraints that are core for the original AE are still essential, although the current system does not yet have the capacity to generate constraints itself as part of its theories. Conditions (7) and (8) ensure that $\mathcal{W}_T$ is not ambiguous. During an alteration from $open(bridge)$ to $half\_open(bridge)$, there is no overlap with a third value, say $closed(bridge)$ in $\mathcal{W}_T$. This does not constrain the overlap with events in $\mathcal{W}_U$ though. An alteration event e with content $bridge(open, closed)$ in $\mathcal{W}_U$ might very well be explained by a set of three events with respective contents $bridge(open, half)$, $bridge(half, half)$ and $bridge(half, closed)$ in $\mathcal{W}_T$. Condition (9) ensures that a cause does not cover a sequence of alterations of which only the last is its effect: if e causes the temperature to rise, it is generally not sensible that when e starts the temperature falls, and the rising of the temperature only starts when e is at its end. Clearly, theories have multiple traces, although the number of events and the set of contents is always the same. As in the previous chapter, we require that at least one of these traces explains the input.

Definition 21. *Theory T* makes sense *of event structure $\mathcal{W}_U$ if it has a trace $\mathcal{W}_T$ such that:*

1. *For every event $e \in W_U$ with content $(o, (p, q))$ there exist events $e_1, e_2 \in W_T$ such that e_1 has content $(o, (p, r))$ for some predicate r, e_2 has content $(o, (s, q))$ for some*

predicate s, $O(e,e_1), O(e,e_2), R_+(e,e_1), R_-(e,e_2)$. Of course, e_1 and e_2 might be equal, in which case $p = s, r = q$.

2. *For every event $e \in W_U$ with content $(o1, o2, (p,q))$ there exist events $e_1, e_2 \in W_T$ such that e_1 has content $(o1, o2, (p,r))$ for some predicate r, e_2 has content $(o1, o2, (s,q))$ for some predicate s, $O(e,e_1), O(e,e_2), R_+(e,e_1), R_-(e,e_2)$. Of course, e_1 and e_2 might be equal, in which case $p = s, r = q$.*

3. *For every event $e \in W_U$ with content $(p, o, (v_1, v_2))$ there exists an event $e_1 \in W_T$ with content $(p, o, (v_3, v_4))$ so that $v_1 <= v_2$ implies $v_3 <= v_4$. $v_1 >= v_2$ implies $v_3 >= v_4$, the interval $[v_1, v_2]$ is enclosed in the interval $[v_3, v_4]$, and $e \preceq e_1$.*

4.5.7 Making sense of event structures: code

The generation of terms, rules, variables and substitutions from a template ϕ is very similar to that of the original AE and is thus not discussed in detail here. An important distinction is of course that rules are now generated in CONTENTLOG instead of DATALOG$^{\ni}$, which requires rewriting of the whole backbone of the AE that facilitates language construction and substitution. The interested reader may look at the code itself. I now however only discuss a few interesting parts of the code that generates a trace $\mathcal{W}_T$ following definition 20. A more extensive overview of this process is given in the appendix. Firstly, generation of events following conditions (1), (2), (3) and (4) is done by means of choice constructs:

```
1 {regular_succession(R,Subs,E1,E2) : pos_concept_event(E2), E1<E2}1:-
        concept_event(E1),
        is_regsucc_rule(R),
        rule_subs(R,Subs),
        eval_body(R,Subs,E1),
        not ending_concept_event(E1).

1 {causes(R,Subs,E1,E2) : pos_concept_event(E2), E1 < E2}1:-
        concept_event(E1),
        is_causal_rule(R),
        rule_subs(R,Subs),
        eval_body(R,Subs,E1),
        not ending_concept_event(E1).

1 {stabilises(s(O,(C1,C2)),E1,E2) : pos_concept_event(E2), E1 < E2}1 :-
        concept_event(E1),
        has_content(E1, s(O,(C1,C2))),
        C1 != C2.

1 {stabilises(s2(O1,O2,(C1,C2)),E1,E2) : pos_concept_event(E2), E1 < E2}1 :-
        concept_event(E1),
        has_content(E1, s2(O1,O2,(C1,C2))),
        C1 != C2.

1 {stabilises(s2v(C,O,(V1,V2)),E1,E2) : pos_concept_event(E2), E1<E2}1 :-
        concept_event(E1),
        has_content(E1, s2v(C,O,(V1,V2))),
```

```
        V1 != V2.
```

Program 4.8: Choosing event tokens for contents that follow from causation, regular succession or stabilisation

Based on these choices, the events are then assigned the contents following from the relevant rule and substitution or the event that is 'stabilised'. The following program then enforces that events are *arbitrarily close* to their successors:

```
future_event(E1,E2) :-
    r_after(E1,E2), not time_overlap(E1,E2).

exist_end_between(E1, E2) :-
    concept_event(E2),not init_event(E2), concept_event(E1),
    concept_event(E3), future_event(E2, E1), r_strictly_before(E1, E3),
    not r_after(E3,E2), not time_overlap(E3, E2).

exist_begin_between(E1, E2) :-
    concept_event(E2), not init_event(E2), concept_event(E1),
    concept_event(E3), future_event(E2,E1), r_strictly_after(E2,E3),
    not r_before(E3,E1), not time_overlap(E3,E1).

boundary_between(E1,E2) :-
    concept_event(E2), not init_event(E2), concept_event(E1),
    future_event(E2,E1), not exist_begin_between(E1,E2),
    not exist_end_between(E1,E2).

:- stabilises(_,E1,E2), not boundary_between(E1,E2).
:- regular_succession(_,_,E1,E2), not boundary_between(E1,E2).
```

Program 4.9: Defining closeness

The remaining conditions from definition 20 mostly correspond to relatively simple constraints. Finally, it must be assured that T makes sense of $\mathcal{W}_U$. As an example, the first condition of definition 21 is partly implemented by a clause of the following form:

```
:- intuit_event(IE), not has_cover(IE).

has_cover(IE) :- has_concept_cover(IE, _).

% Content_begin_covered and content_end_covered
% are true of the beginning or respectively ending contents are equal.
has_concept_cover(IE, (CB,CE)) :-
    intuit_event(IE), concept_event(CB),
    concept_event(CE),
    content_begin_covered(IE, CB),
    content_end_covered(IE, CE),
    time_overlap(IE, CB),
    time_ovelap(IE, CE),
    r_after(IE, CB), r_before(IE, CE).
```

Program 4.10: Determining when an event with content of the form $(o, (p, q))$ is conceptually covered

I have also left out the translation from contents of the form (c, o, up) to contents of the form $(c, o, (v_1, v_2))$, which is a guess over the domain of numerical constants.

4.6 Example behavior

Example 3. The FAE is given a template $(\phi, N_i, N_e, N_{RS}, N_{CR})$, where ϕ is a type signature, and the remaining arguments now determine the maximum numbers of init atoms, events in the trace of a constructed theory, *regular successions* and *causal rules*. As a starter, I apply the FAE to one of the simple example problems for the original AE:

$$
\begin{aligned}
S_1 &= \{\} \\
S_2 &= \{off(sensor\_a), p_a(sensor\_b)\} \\
S_3 &= \{on(sensor\_a), p_a(sensor\_b)\} \\
S_4 &= \{off(sensor\_a), p_a(sensor\_b)\} \\
S_5 &= \{\} \\
S_6 &= \{off(sensor\_a), p_a(sensor\_b)\}
\end{aligned}
$$

But I add one additional sensation: $S_7 = \{off(sensor\_a)\}$. The provided template is the following:

$$
\Phi = \left\{\begin{aligned}
&T = \{sensor\_1, sensor\_2\}, \\
&O = \{sensor\_a : sensor\_1, sensor\_b : sensor\_2\} \\
&P = \{off(s\_1), on(s\_1), p_a(s\_2), \\
&\qquad p_b(s\_2), p_c(s\_2)\} \\
&V = \{s\_1 : sensor\_1, s\_2 : sensor\_2\}
\end{aligned}\right\}
\begin{aligned}
N_i &= 2 \\
N_e &= 16 \\
N_{RS} &= 2 \\
N_{CR} &= 0
\end{aligned}
$$

And the constraints in the input:

$$
C = \left\{\begin{aligned}
&\forall x : \ sensor\_1 \ (on(x) \ XOR \ off(x)) \\
&\forall y : \ sensor\_2 \ (p_a(y) \ XOR \ p_b(y) \ XOR \ p_c(y))
\end{aligned}\right\}
$$

The original AE solves the task without S_7 with a theory where $I = \{on(sensor\_a), p_a(sensor\_b)\}$ and R contains two causal rules: $\{on(sensor\_a) \ni off(sensor\_a)\}$ and $\{off(sensor\_a) \ni on(sensor\_a)\}$. The FAE finds a similar theory, but now also constructs event structure $\mathcal{W}_U$ and trace $\mathcal{W}_T$:

$$
I = \left\{off(sensor\_a), p\_a(sensor\_b)\right\} \ R = \left\{\begin{aligned}
&(s\_1, (on, on) \rightarrow_{RS} (s\_1, (on, off)) \\
&(s\_1, (off, off)) \rightarrow_{RS} (s\_1, (off, on))
\end{aligned}\right\}
$$

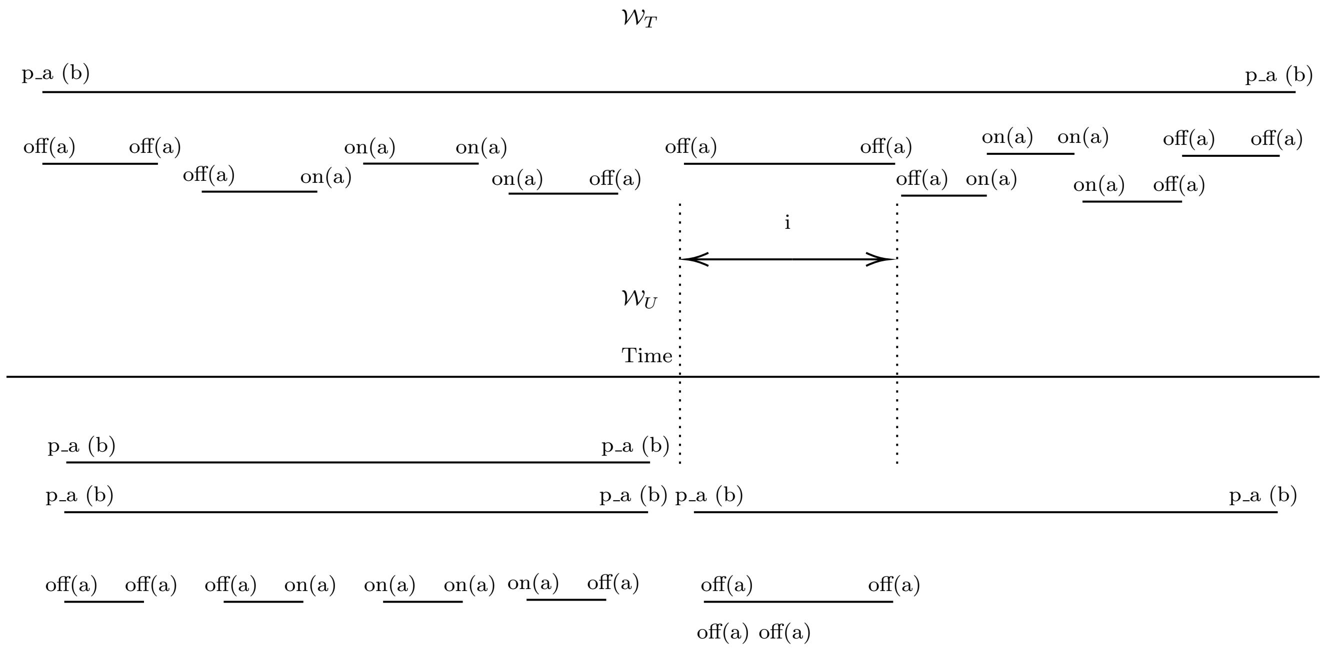


Figure 4.4: $\mathcal{W}_T$ and $\mathcal{W}_U$ for example 3, i denotes a minimal interval including two sensations that are successive in the input ordering.

△

Note that $\mathcal{W}_U$ and $\mathcal{W}_T$ in fact together constitute a single event structure, in which every intuition event in $\mathcal{W}_U$ is covered or 'explained' by an event in $\mathcal{W}_T$ that follows from the conceptual theory. $\mathcal{W}_T$ also continues beyond the time of input in $\mathcal{W}_{\mathcal{U}}$, providing predictions of future sensations from the conceptual structure.

Let's recall how $\mathcal{W}_U$ and $\mathcal{W}_T$ have been constructed. For reasons of clarity I use a slightly procedural terminology to explain the program as if it operates step by step, but it is important to note that the whole system program is in fact solved as a single set of logical constraints. The FAE constructs 6 manifolds from the input, each of which contains the *stable* events associated with the sensations in the corresponding input position. The events in each of these manifolds must overlap with each other. Now a sequence of 'merged' manifolds is constructed to which there exist embeddings. For instance, manifolds $(1,0)$ and $(2,0)$ are embedded into manifold $(1,1)$, which is a copy of $(2,0)$ since $(1,0)$ only contains the temporal cover. Now $(1,1)$ and $(3,0)$ are embedded into a manifold $(1,2)$. Since the FAE then finds that sensation $off(sensor_a)$ is followed in the input ordering by $on(sensor_a)$, and since it knows by the constraints in C that an object cannot be both on and off at the same time, it deduces that the stable event corresponding to $on(sensor_a)$ must end strictly before the stable event corresponding to $off(sensor_a)$ starts. The space of possible temporal structures is constrained by the contents of the constructed events. Furthermore, the FAE deduces that there must exist an *alteration event* in W_U with content $(sensor_a, (off, on))$ between the two stable events, representing s change from off to on. Note that at this point, $(1,2)$ also contains two events with content $(sensor_b, (p_a, p_a))$, but the precise temporal structure of these events, e.g. whether or they cover the alteration from off to on or each other, has not been determined. Iterating this process then produces the final manifold $(1, 5)$ which contains all events in $\mathcal{W}_U$ and their contents. Again, the precise temporal structure of $\mathcal{W}_U$ has to a large extent not been determined at this point. For instance, whether the two last sensations of $off(sensor_a)$ should be represented as sensations of the same stable state, as is done in $\mathcal{W}_U$, or as two distinct stable states between which other alterations might have occurred, is decided in the process of *making sense* through $\mathcal{W}_T$.

Now for constructing the trace $\mathcal{W}_T$, the FAE must find suitable init atoms and rules that can be constructed from ϕ, the resulting events must then cover all events in $\mathcal{W}_U$. The init atoms mus correspond with the 'earliest' events in $\mathcal{W}_T$. The FAE decides on $off(sensor_a)$ as an init atom, and uses it to cover the first event with content $(sensor_a, (off, off))$. Other choices are also possible. For instance, if it had resulted in a simpler conceptual explanation the FAE could have hypothesized that $sensor_a$ was first on, and a sequence of changes had happened before the first sensation $off(sensor_a)$. Now the FAE hypothesis that $p_a(sensor_b)$ is true initially and does not change, producing a single conceptual event with content $(sensor_b, (p_a, p_a))$ that explains all corresponding sensations. The FAE also finds two rules of regular succession, explaining how stable events are regularly succeeded by alteration events. These alteration events must in turn always be followed by stable events following definition 20. Note that each of the associated pairs of concept events are arbitrarily close and hence separated by unique boundaries; no event begins or ends in the space between the pair. The concept events then 'explain' the sensations through temporal covering. Note that the two last events in $\mathcal{W}_U$ with content $(sensor_a, (off, off))$ are interpreted as a single stable event, which allows their covering under a single stable concept event with the same content. Furthermore, they also fall within the same interval i. The succession in the input ordering is thus not translated into an objective succession via the covering in $\mathcal{W}_T$ nor in a succession of moments in the manifold of intuition $\mathcal{W}_U$).

I now provide a more complicated example, showing how the FAE can model processes of efficient causation. The FAE is asked to derive the mechanics of a basic coffee maker. The relevant objects are 'water' and 'maker'. The intended interpretation of the sensory sequence is that an agent switches the coffee maker from 'off' to 'on', after which the water moves through different stages and different temperatures until coffee is produced. The action of activating the coffee maker needs not be explained, so the FAE is given an 'exogenous action' $(maker, (off, on))$ between input positions 1 and 2, that gives rises to an event in $\mathcal{W}_T$, without following from the theory. However, the precondition of this action (the maker being off) does require an explanation.

Example 4.

$$
\begin{aligned}
S_1 &= \{\} \\
S_2 &= \{basin(water), temp(water, 2)\} \\
S_3 &= \{temp(water, 6)\} \\
S_4 &= \{temp(water, 10)\} \\
S_5 &= \{vape(water)\} \\
S_6 &= \{temp(water, 5)\} \\
S_7 &= \{temp(water, 2)\} \\
S_8 &= \{filter(water)\} \\
S_9 &= \{on(maker), coffee(water)\} \\
S_{10} &= \{off(maker)\}
\end{aligned}
$$

The provided template is the following:

$$
\Phi = \left\{
\begin{aligned}
&T = \{liquid, machine\}, \\
&O = \{water : liquid, maker : machine\} \\
&P = \{basin(y), vape(y), filter(y), \\
&\qquad coffee(y), on(x), off(x)\} \cup \\
&\qquad \{temp(y, i) : 0 \leq i \leq 10)\} \\
&V = \{x : machine, y : liquid\}
\end{aligned}
\right\}
\begin{aligned}
N_i &= 3 \\
N_e &= 15 \\
N_{RS} &= 1 \\
N_{CR} &= 5
\end{aligned}
$$

And the constraints in the input:

$$
C = \left\{
\begin{aligned}
&\forall x : \ liquid \ (basin(x) \ XOR \ vape(x) \ XOR \ filter(x) \ XOR \ coffee(x)) \\
&\forall y : \ maker \ (on(y) \ XOR \ off(y))
\end{aligned}
\right\}
$$

The FAE finds the following theory and event structures $\mathcal{W}_U, \mathcal{W}_T$:

$$
I = \Big\{off(sensor\_a), basin(water), temp(water, 2)\Big\}
$$

$$
\text{R} = \left\{
\begin{aligned}
&(y, (on, on) >>_{CR} (temp, x, up) \\
&(temp, x, up) >>_{CR} (x, (basin, vape)) \\
&(temp, x, (10, 10)) >>_{CR} (x, (vape, filter)) \\
&(x, (filter, filter)) >>_{CR} (temp, x, down) \\
&(temp, x, down) \rightarrow_{RS} (x, (filter, coffee)) \\
&(x, (coffee, coffee)) >>_{CR} (y, (on, off))
\end{aligned}
\right\}
$$

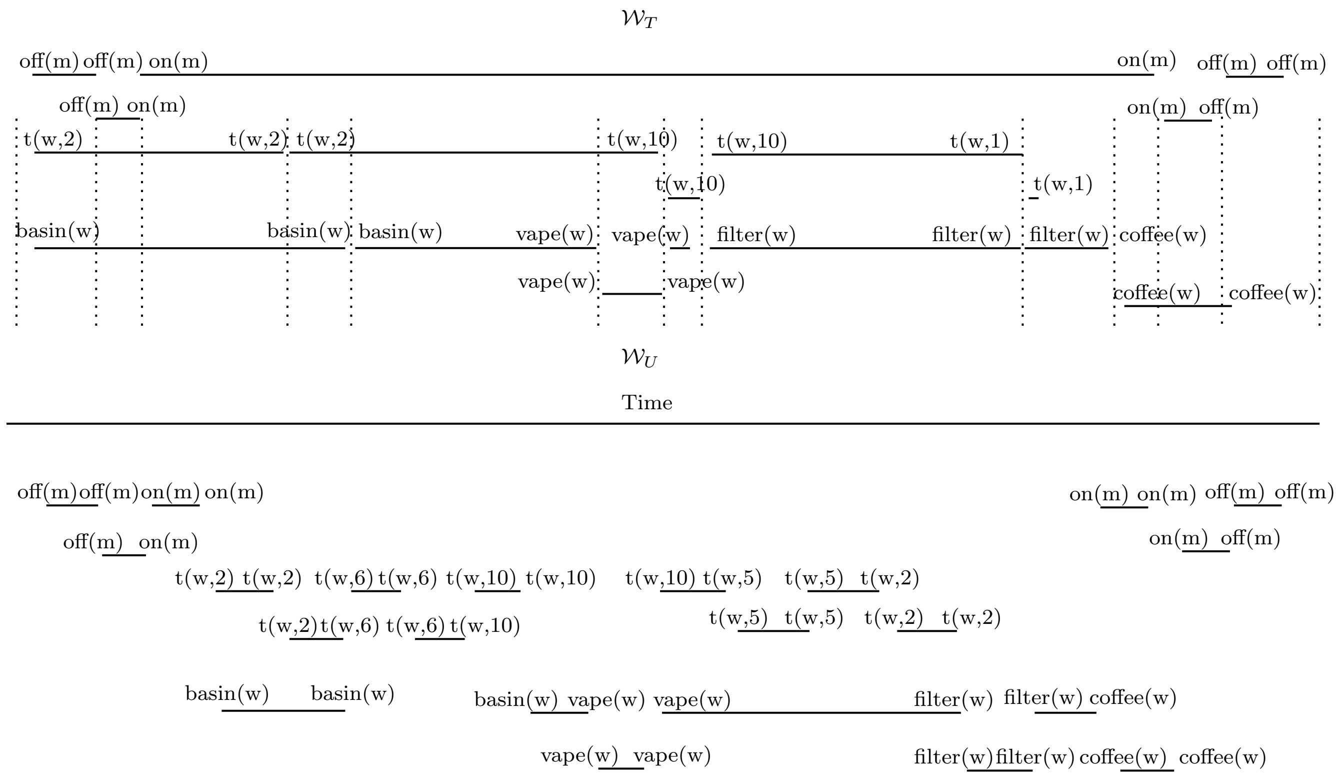


Figure 4.5: Event structures for example 4, where w and m represent the objects 'water' and 'maker' respectively. All boundaries in $\mathcal{W}_T$ are denoted by dotted lines, between which the intervals of $\mathcal{W}_T$ are enclosed.

△

Again let's go trough how $\mathcal{W}_U$ and $\mathcal{W}_T$ have been constructed. All events in $\mathcal{W}_U$ are either direct results of sensations, or produced by the FAE as transitions between sensations. Note that each sensed temperature brings forth a distinct stable event. Again, some temporal relations have been determined by the input. For instance, the event with content $(water, (basin, basin))$ must overlap with the first event with content $(temp, water, (2, 2))$ since both come from the same position in the input ordering. The event with content $(water, (basin, basin))$ must also start before the event with content $(temp, water, (6, 6))$ ends, since the latter is sensed later in the input ordering, but the larger part of temporal relations must be chosen by the FAE to allow for conceptual interpretation. The exogenous action $(maker, (off, on))$ between positions 1 and 2 gives rise to three events in $\mathcal{W}_U$. If the maker is changed 'exogenously' from off to on, it must have been off before the action, and it must be on after the action, which primes two stable events. Note that only the exogenous action itself however exists both in $\mathcal{W}_U$ and $\mathcal{W}_T$ since it need not be explained conceptually. The stable events before and after the exogenous action do need a conceptual explanation, and are thus covered by distinct events in $\mathcal{W}_T$. The init event with content $(maker, (off, off))$ is produced by an init atom in the theory, and the event with content $(maker, (on, on))$ is the stable event resulting from the alteration $(maker, (off, on))$. Note that in the trace $\mathcal{W}_T$, the intervals between boundaries are very different from the input ordering. For instance, a single interval enclosing $(temp, water, (2, 10))$ also contains multiple successive sensations of temperature. Note also that the stable event with content $(maker, (on, on))$ that results from the exogenous action in $\mathcal{W}_T$ is protracted to cover two sensations. This allows this event to function as cause for the rising of the temperature following the first rule since it covers the event in $\mathcal{W}_T$ with content $(temp, water, (2, 10))$. Finally, we can see that every event in $\mathcal{W}_T$ is *arbitrarily close* to and hence separated by a unique boundary from a later event so that the ending content of the former is the starting content of the latter, with the exception of the events in the last interval. This ensures that $\mathcal{W}_T$ produces a complete explanation, there are no 'clefts' in time in which the $\mathcal{W}_T$ does not determine the state of one of the objects.

I have included one *regular succession* in this example to show how both types of rules can be applied in a single theory, but the alteration from 'filter' to 'coffee' could also have been explained by a causal rule, such as $(temp, x, (1, 1)) >>_{CR} (x, (filter, coffee))$. The reader might note that the rules do not specify any binary relation connecting x and y. The first rule thus strictly speaking states universally that if *any* machine is on, this causes the temperature of *any* liquid to rise, but the interpretation is sensible due to the domain restrictions for x and y. I have not implemented the possibility for including multiple atoms in a body, so that rules of the form $(y, (on, on)) \wedge (x, y, (in, in)) >>_{CR} (temp, x, up))$ that restrict their applicability using binary relations can not yet be constructed. The FAE does have the capacity to reason about 'intensities' (numerical values) both on the level of specific values and on the level of general directions (up and down).

Now finally, I provide an example where input is given as partial event structure

Example 5. I provide a partial event structure as input. Unfortunately the representation of a partial event structure as a set of formulas is quite cumbersome, but a representation in the usual manner would be incorrect since this would suggest complete knowledge of the temporal relations.

O	$((\mathit{speed}, \mathit{cart\_a}, 0), 1), ((\mathit{push}, \mathit{bot}, \mathit{cart\_a}), 1)$
R_+	$((\mathit{speed}, \mathit{cart\_a}, 0), 1), ((\mathit{push}, \mathit{bot}, \mathit{cart\_a}), 1)$
R_-	$((\mathit{speed}, \mathit{cart\_a}, 0), 1), ((\mathit{push}, \mathit{bot}, \mathit{cart\_a}), 1)$
R_-	$((\mathit{speed}, \mathit{cart\_a}, 0), 1), ((\mathit{speed}, \mathit{cart\_a}, 0), 2)$
R_+	$((\mathit{push}, \mathit{bot}, \mathit{cart\_a}), 1), ((\mathit{speed}, \mathit{cart\_a}, 0), 1)$
R_-	$((\mathit{push}, \mathit{bot}, \mathit{cart\_a}), 1), ((\mathit{speed}, \mathit{cart\_a}, 5), 1)$
R_-	$((\mathit{push}, \mathit{bot}, \mathit{cart\_a}), 1), ((\mathit{release}, \mathit{bot}, \mathit{cart\_a}), 1)$
R_-	$((\mathit{speed}, \mathit{cart\_a}, 5), 1), ((\mathit{push}, \mathit{bot}, \mathit{cart\_a}), 1)$
O	$((\mathit{speed}, \mathit{cart\_a}, 5), 1), ((\mathit{push}, \mathit{bot}, \mathit{cart\_a}), 1)$
R_-	$((\mathit{speed}, \mathit{cart\_a}, 5), 1), ((\mathit{release}, \mathit{bot}, \mathit{cart\_a}), 1)$
R_-	$((\mathit{speed}, \mathit{cart\_a}, 5), 1), ((\mathit{speed}, \mathit{cart\_a}, 0), 2)$
$\preceq$	$((\mathit{speed}, \mathit{cart\_a}, 5), 1), ((\mathit{release}, \mathit{bot}, \mathit{cart\_b}), 1)$
R_+	$((\mathit{release}, \mathit{bot}, \mathit{cart\_a}), 1), ((\mathit{push}, \mathit{bot}, \mathit{cart\_a}), 1)$
R_-	$((\mathit{release}, \mathit{bot}, \mathit{cart\_a}), 1), ((\mathit{speed}, \mathit{cart\_a}, 0), 2)$
O	$((\mathit{speed}, \mathit{cart\_a}, 0), 2), ((\mathit{release}, \mathit{bot}, \mathit{cart\_a}), 1)$
$\preceq$	$((\mathit{speed}, \mathit{cart\_a}, 0), 2), ((\mathit{release}, \mathit{bot}, \mathit{cart\_a}), 1)$
R_+	$((\mathit{speed}, \mathit{cart\_a}, 0), 2), ((\mathit{release}, \mathit{bot}, \mathit{cart\_a}), 1)$
R_+	$((\mathit{speed}, \mathit{cart\_a}, 0), 2), ((\mathit{speed}, \mathit{cart\_a}, 5), 1)$
R_+	$((\mathit{speed}, \mathit{cart\_a}, 0), 2), ((\mathit{speed}, \mathit{cart\_a}, 0), 1)$
$\preceq$	$((\mathit{release}, \mathit{bot}, \mathit{cart\_b}), 1), ((\mathit{push}, \mathit{bot}, \mathit{cart\_a}), 1)$
$\preceq$	$((\mathit{release}, \mathit{bot}, \mathit{cart\_b}), 1), ((\mathit{speed}, \mathit{cart\_a}, 0), 1)$
$\preceq$	$((\mathit{release}, \mathit{bot}, \mathit{cart\_a}), 1), ((\mathit{release}, \mathit{bot}, \mathit{cart\_b}), 1)$
$\preceq$	$((\mathit{release}, \mathit{bot}, \mathit{cart\_b}), 1), ((\mathit{speed}, \mathit{cart\_a}, 0), 2)$

Table 4.1: Input for example 5, missing atomic relations are interpreted as undetermined: the FAE may decide whether they are true or false.

Note that this only determines a set of positive atomic formulas. If an atomic formula $R_+(a, b)$ is not part of the input, the FAE must decide whether $R_+(a, b)$ or $\neg R_+(a, b)$ is the case. Note further that the input only specifies temporal relations between *sensations* and not *events*. As in the previous examples, the system constructs events with a stable content from these sensations, before introducing *alteration events* to explain changing sensations. The system is further provided with the following template and input constraint:

$$\Phi = \left\{ \begin{aligned} &T = \{agent, cart\}, \\ &O = \{bot : agent, cart_a : cart, cart_b : cart\} \\ &P = \{push(x, y), release(x, y)\} \cup \\ &\qquad \{speed(y, i) : 0 \leq i \leq 10)\} \\ &V = \{x : agent, y : cart\} \end{aligned} \right\} \begin{aligned} N_i &= 4 \\ N_e &= 12 \\ N_{RS} &= 2 \\ N_{CR} &= 2 \end{aligned}$$

$$C = \big\{\forall x, y : \ agent, cart \ (push(x, y) \ XOR \ release(x, y))\big\}$$

The FAE now first constructs an event structure $\mathcal{W}_I$ from the input sensations, extending the partial event structure from the input to an event structure, and identifying intervals.

$\mathcal{W}_I$

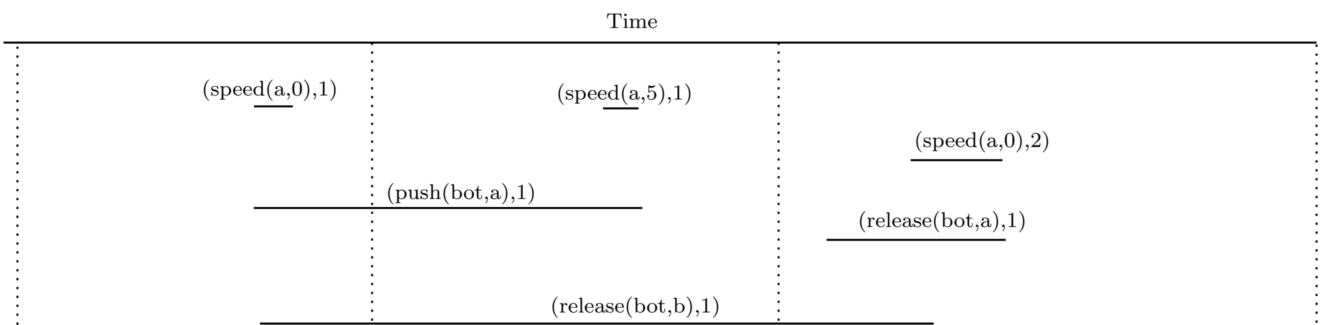


Figure 4.6: Event structure that extends the input, a, b denote *cart_a*, *cart_b*. Boundaries are represented as dotted lines.

Note that each event is now labelled by a sensation instead of a content, and note that there are two distinct sensations $(speed, cart\_a, 0)$, distinguished by the token-indices 1 and 2. The FAE has added atomic formulas such as $r_+(\ ((speed\,(cart\_a, 5)), 1), ((push, (bot, cart\_a)), 1))$ and $r_-(\ ((speed\,(cart\_a, 0)), 1), ((speed\,(cart\_a, 5)), 1))$, but most missing atomic formulas in the input have been decided negatively. The reader may note however that a rather large part of $\mathcal{W}_I$ has been specified in the input. Unfortunately, trimming the input further requires a working memory above the 16GB of a standard laptop, since the FAE must choose between all possible extensions of the input on grounds of optimality on the level of $\mathcal{W}_T$ via $\mathcal{W}_U$. Now finally, the figure above shows that the FAE distinguishes 3 'times' as intervals in the input. The sensations now give rise to stable events that are placed in the manifolds $(x, 0)$ associated with the intervals, i.e. $x \in \{1, 2, 3\}$. The temporal relations between sensations are exactly the temporal relations between the associated stable events in $\mathcal{W}_U$, but of course

$\mathcal{W}_U$ also contains the merged events. The resulting theory and event structures are then as follows:

$$I = \{push(bot, cart\_a), push(bot, cart\_b), speed(cart\_a, 0), speed(cart\_b, 2) \quad \}$$

$$\text{R} = \begin{Bmatrix} (x, y, (push, push)) >\!\!>_{CR} (speed, y, up) \\ (x, y, (push, release)) >\!\!>_{CR} (speed, y, down) \\ (x, y, (push, push)) \rightarrow_{RS} (x, y, (push, release)) \end{Bmatrix}$$

△

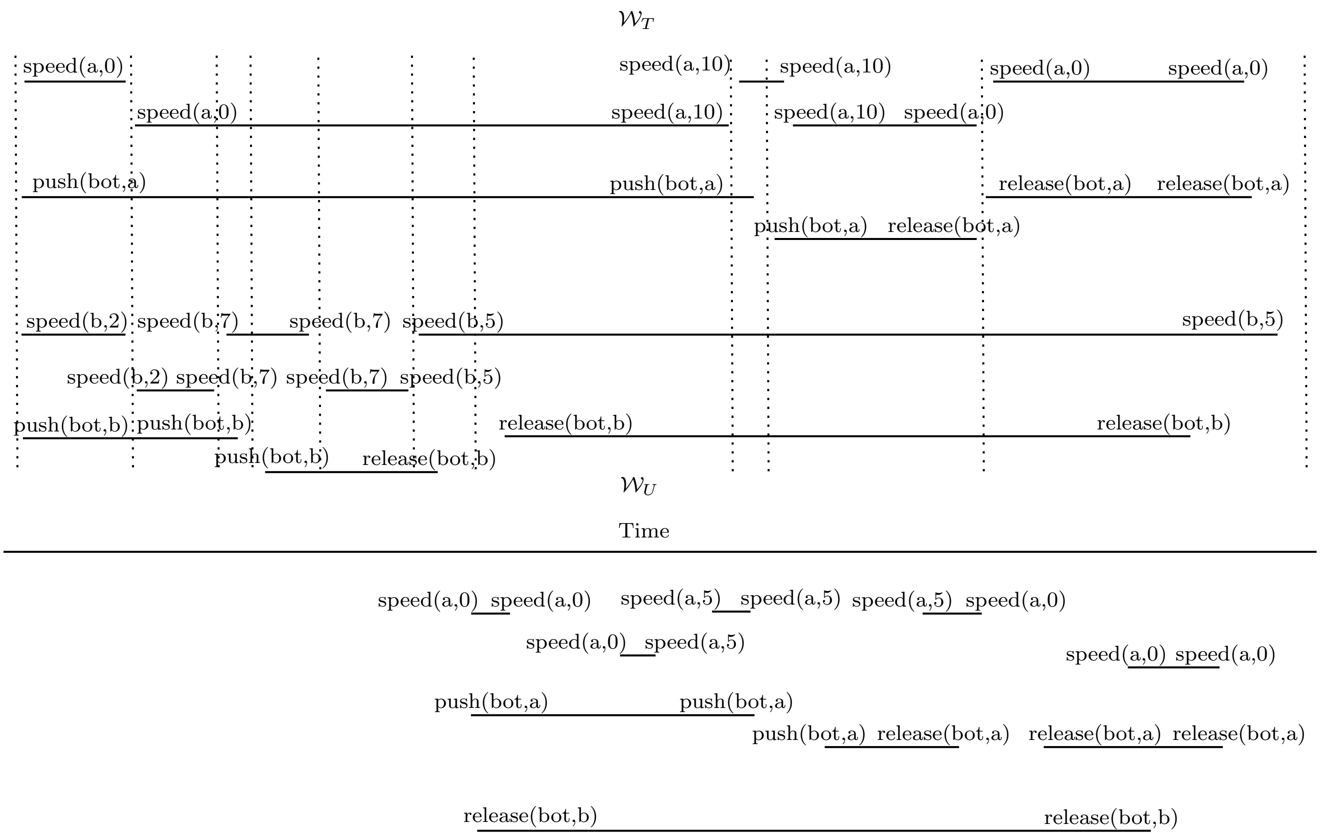


Figure 4.7: Event structures for example 5, a, b represent $cart\_a, cart\_b$ and boundaries are denoted by dotted lines.

Here we now see many benefits of the expressive power of the FAE at work. Clearly the 3 intervals in the input ordering are only the starting point of the final temporal representation, comprising a total of 10 intervals. Times have been added in the clefts between times. For instance, the sensations corresponding with the second and third interval in $\mathcal{W}_I$ are placed in the 8th and 10th interval in $\mathcal{W}_U$; an additional moment in time has been constructed in between. We can also see how causation has been separated from a succession in causal moments. The same causal rule $(x, y, (push, push) >>_{CR} (speed, y, up)$ explains the rising of the speed of *cart_a* over the course of 6 temporal intervals and the rising of the speed of *cart_b* over the course of 1 interval. Finally, the FAE shows its predictive value by filling in speeds for *cart_b* matching the expected pushing and releasing by the 'bot'.

Conclusion

This then concludes the fourth chapter. The FAE structures intuition in time; a process that is a determination of intuition by the understanding. Its construction of $\mathcal{W}_U$ represents the syntheses of apprehension and reproduction, but in its pure application the same construction represents the *figurative synthesis* that brings forth a unified *formal intuition*. This *formal intuition* is a pre-conceptual unity, since it is only the manifold of intuition that must be conceptually unified through $\mathcal{W}_T$. Still, $\mathcal{W}_U$ is a determination of sensibility by the understanding, since the conceptual theory structures $\mathcal{W}_U$ and since the relations from which $\mathcal{W}_U$ is built are functions under the categories. This is then quite a satisfying representation of the interpretation given by Beatrice Longuenesse and introduced in chapter 2 that the *figurative synthesis* underlies the threefold synthesis of apprehension, reproduction and representation that constitutes experience. I now turn towards the form of outer intuition: space.

Chapter 5

Space: Theory

5.1 Introduction

We now turn to the spatial component of figurative synthesis. Kant states that the succession of moments in time that generates a manifold of intuition is the result of "*motion*, as action of the subject". This very motion Kant however also characterises as "synthesis of the manifold in space" (B155), so that we can only represent time "under image of a line, insofar as we draw it" (B156). Synthesis in space and time thus cannot be understood as distinct processes, which explains the term *spatiotemporal synthesis* (i.e., *figurative synthesis*). If we want to take *figurative synthesis* seriously, we must thus apply it to outer sense as well as inner sense. Recalling our findings in chapter 2, the spatial structure must be an "infinite magnitude" (B40) (in the potential sense discussed in the previous chapter), as well as "essentially single". As is the case for times, spaces can only be represented if the *a priori* representation of space as a unique whole is their ground (A24). Furthermore, Kant understood space as *a priori* ground for three-dimensional geometry.

In this chapter, I add a spatial construction to the implementation of figurative synthesis introduced in the previous chapter. Since *figurative synthesis* is a determination of sensibility by the understanding, this construction must again be *qualitative*: spaces are logical models of the Regional Connection Calculus (Cohn, Bennett, Gooday, & Gotts, 1997). Again a topological interpretation is of use to represent the Kantian desiderata in relation to space as *formal intuition*. I introduce atomic regions as a representation of Kants 'extensive magnitude', provide them with a directional orientation and analyse space as connection graph. This allows an analysis on the levels of dimensionality and global configuration. The implementation of this framework is then the topic of the next chapter.

5.2 Regional Connection Calculus

The Regional Connection Calculus is a qualitative spatial resoning structure that was originally developed at the University of Leeds in the 1990s. The pioneering book was written by Randell, Cui and Cohn (1992), from which the acronym RCC originates. The authors opted for the usage of regions instead of points as reasoning primitives on phenomenological grounds: even if we talk about points in daily life (e.g., the point of a pencil), we usually mean regions and not mathematical points. Furthermore, they argued that representation of space as an infinite set of points is not phenomenologically feasible; an argument that shows interesting similarities with Kant's conception of potential infinity. The intended meaning of 'region' in the calculus is a non-null spatially extended entity. All regions in a model have the same dimension, and no regions have 'mixed' dimension, i.e., a 2D plane with a 3D bulb sticking out. This condition is defined as space being *regular*. Regions with holes or disconnected parts are in general allowed. The primitive relation is that of a reflexive and symmetric binary *connection* C between regions:

$$\forall x\, C(x, x)$$
$$\forall x, y\, C(x, y) \rightarrow C(y, x)$$

The calculus includes a constant u, representing the *universal region.* Hence we have

$$\forall x\, C(x, u)$$

At this point it is important to note that both spatial and temporal axioms include universal quantification, but not necessarily over the same domain. Universal quantification in the temporal axioms ranges over a temporal domain and the universal quantification above ranges over a spatial domain. However, for the FAE these two domains are almost identical. The temporal domain is a strict subset of the spatial domain, and the only spatial elements that do not feature in time are the *sub-atomic* regions that are introduced later in this chapter. This then ensures that the same input sensations are structured in both space and time. Now from the connection relation C, all other relations between regions are defined as follows (Cohn et al., 1997):

1. Disconnected: $DC(x, y) := \neg C(x, y)$
2. Part $P(x, y) := \forall z\, C(z, x) \rightarrow C(z, y)$
3. Proper part $PP(x, y) := P(x, y) \wedge \neg P(y, x)$
4. Identity $EQ(x, y) := P(x, y) \wedge P(y, x)$
5. Spatial overlap $SO(x, y) := \exists z\, (P(z, x) \wedge P(z, y))$
6. Discrete from $DR(x, y) := \neg SO(x, y)$
7. Partial overlap
   $PO(x, y) := SO(x, y) \wedge \neg P(x, y) \wedge \neg P(y, x)$
8. Externally connected:
   $EC(x, y) := C(x, y) \wedge \neg SO(x, y)$
9. Tangenial proper part:
   $TPP(x, y) := PP(x, y) \wedge \exists z\, (EC(z, x) \wedge EC(z, y))$
10. Nontangenial proper part:
    $NTPP(x, y) := PP(x, y) \wedge \neg\exists z\, (EC(z, x) \wedge EC(z, y))$
11. Tangenial proper part inverse:
    $TPPi(x, y) := TPP(y, x)$

12. Nontangenial proper part inverse:
   $NTPPi(x, y) := NTPP(y, x)$

A spatial structure $\mathcal{S} = (S, C)$ now contains a finite set of regions S on which the connection relation C is defined, so that all other RCC relations are also determined. Generally, RCC further contains the (partial) boolean functions *sum*, *product* (intersection) and *complement*. As in the previous chapter, I do not use functions unrestrictedly, since they result in a combinatorial explosion of regions. In the case of space the functions even render the system undecidable (Dornheim, 1998). The downside of this restriction is that it prevents us from expressing the connectedness of regions in the usual manner as $\forall yz : (EQ(x, sum(y, z)) \rightarrow C(y, z))$, but a solution for this issue will be introduced later. The relations $\{DC, EC, PO, EQ, TPP, NTPP, TPPi$ and $NTPPi\}$ are jointly exhaustive and pairwise disjoint (JEPD), and RCC reduced to these 8 relations is named RCC-8. This restriction of RCC has given rise to efficient reasoning methods by means *composition tables* (see e.g., (Li & Ying, 2003)). It has been shown that any consistent set of RCC-8 relations has a model in which the regions are self-connected subsets of a space $\mathbb{R}^n$ for $n \geq 3$, while an interpretation over self-connected regions in two-dimensional space might not be possible for sets of RCC-8 and hence RCC relations (Grigni, Papadias, & Papadimitriou, 1995). Dropping the connectedness constraint, model existence has been shown for any of $\mathbb{R}^n$ with $n \geq 1$, although satisfibility for RCC-8 is in general an NP complete problem. Since our aim here is *model construction* rather than *reasoning*, construction from C is more efficient than reasoning by composition tables. I thus make use of full RCC. If extensionality holds for an RCC model: $EQ(x, y) \rightarrow x = y$, it is termed *strict*. Strict RCC models are equivalent to Boolean connection algebras (see e.g., (Ligozat, 2013, Ch.10)). In the strict case, it can be shown that for any element a in the model except u, and any RCC relation R, there exists an element x in the model such that $R(a, x)$ holds. The following axiom that prevents finite models then follows $\forall x \exists y : NTPP(y, x)$. As in the previous chapter, I do not apply extensionality however, 'intuitions' with different contents can be assigned the same space. This allows the FAE to construct finite spatial structures, following Kant's conception of *potential infinity* as progression in intuition.

5.3 Topological interpretation

Although RCC was intended as a point-free approach towards spatial reasoning, it has intuitive interpretations in point-set topology. Usually, for sets $A, B \in \mathcal{P}(X)$, $C(A, B)$ holds iff the closures of A and B have a non-empty intersection, where the closure $cl(A)$ of A is the smallest closed set containing A. The interior $int(A)$ is conversely the largest open set contained in A. Name a topological space regular if for any pair (x, Y), where x is a point and Y is a closed set such that $x \notin Y$, there exist two disjoint open sets containing x and Y, respectively. Further, let a regular closed set be a closed set such that $cl(int(A)) = A$, and a regular open set $A \in \tau(X)$ such that $int(cl(A)) = A$. Then, it has been shown that if X is a non-empty connected and regular topological space, an RCC model is formed by either its non-empty regular closed sets or its non-empty regular open sets (Ligozat, 2013, Ch.10). A difficulty is posed however by the functions that are generally included in RCC. For instance, the complement of an open set is a closed set and the union of two regular open sets is not necessarily regular. One solution for this issue is to regard regions as equivalence classes of sets with the same closures (Cohn et al., 1997).

In this project I however interpret RCC using point-free models, so that connectedness of space cannot be evaluated as in standard topology. As said, I also do not make use of the unrestricted *sum* function that is often applied to ensure connectedness. Still, we would like to implement Kant's claim that parts of space are spaces that can only be represented as limitations of space as connected whole. This can be achieved by reasoning about sets of regions and evaluating the topology of this higher-level structure. Once again, we can apply the Alexandroff topology, noting that the inverse part relation P^{-1} forms a pre-order. For spatial structure $\mathcal{S}$ consider the Alexandroff topology $\mathcal{A}_{P^{-1}}$ as defined in the previous chapter, where open sets are downsets of P. Then the following holds, since every non-empty closed set of $\mathcal{A}_{P^{-1}}$ contains the universal region u:

Proposition 4. *For any spatial structure $\mathcal{S}$, the topology $\mathcal{A}_{P^{-1}}$ is ultra-connected.*

Now identifying parts of space (i.e. spaces) with a P downset generated from region $x \in S$, we can easily see that space as a whole cannot be generated from its parts, but spaces are only limitations of the whole space .

Proposition 5. *Let $\mathcal{S}$ be a spatial structure with the $\mathcal{A}_{P^{-1}}$ topology. Take open sets $U, V \in \mathcal{A}_{P^{-1}}$ such that $U \cup V = S$. Then $U = S$ or $V = S$.*

Proof. Let $A, B \in \mathcal{A}_{P^{-1}}$ such that $A \cup B = S$. We have $u \in A$ or $u \in B$. W.l.o.g. suppose the former, then $\forall x \in S : C(x, u)$ so that $\forall x, z \in S : (C(x, z) \rightarrow C(x, u))$ so that $\forall z \in S : P(z, u)$. Thus $A = S$. □

This ultra-connectedness on the level of *spaces* does not however ensure that individual regions make up a single connected whole. This is addressed in the next section.

5.4 Atomic regions and extensive magnitude

In the second chapter I noted that space and time are images of Quantity under the schema of number: addition of the homogeneous. While this does not imply that space and time can be reduced to mere counting structures, it does point towards fundamental parallels that Kant saw between counting and sensing in space and time. Indeed, for Kant the representation of a number (unit) required both an *apprehensive act* which produces a multiplicity, and a *comprehensive act* which synthesizes the multiplicity in a unified representation. The latter requires simultaneous representation by "placing them beside one another in space" (R6314). This explains why Kant insisted that time can only be represented by "drawing a straight line" (B154), and points towards the importance of a metric or unit in all synthesis in space and time, rendering appearances *magnitudes*. In fact, appearances in space and time are *extensive magnitudes* "in which the representation of the parts makes possible the representation of the whole" (B203). Note the subtle tension of this bottom-up framework with the dependency on *formal intuition* stressed before: synthesis operates by composition of units, but can only do so by virtue of the unity that is its transcendental ground.

I represent extensive nature of space by means of *atomic regions* $A \subset S$ for which the following holds:

Axiom 1. *Let* $A_1, A_2 \in A \subseteq S$. *Then* $SO(A_1, A_2) \rightarrow EQ(A_1, A_2)$.

The FAE now assigns input 'intuitions' to temporal events as well as atomic regions. I further require that all non-atomic regions are sums over a path-connected set of atomic regions, with the exception of *sub-atomic regions* which I introduce later. I define a new function of 'generalized sum' from path connected sets of atomic regions to regions:

Definition 22. *Let* $A_1 = \{a_1, \ldots, a_n\} \subseteq A \subseteq S$ *with* $A_1 \neq \emptyset$, *and let* $C^*_{A_1}$ *be the transitive closure of the restriction of* C *to* A_1. *Then, if for every* $a_1, a_2 \in A_1$ *it holds that* $C^*_{A_1}(a_1, a_2)$ *we have* $sum(A_1) = sum(a_1, sum(a_1, ...sum(a_{n-1}, a_n)..))$, *where* $x = sum(a_1, a_2)$ *iff* $\forall z \in S : C(z, x) \leftrightarrow C(z, a_1) \vee C(z, a_2)$, *as usual in full RCC.*

Axiom 2. *Let* $x \in S \setminus A$. *If there does not exist* $a \in A$ *such that* $PP(x, a)$, *then there exists* $A_1 \subseteq A$ *such that* $x = sum(A_1)$.

This then resolves the undefinability of connected regions addressed in the previous section. Regions are either atomic or sums over path-connected sets of atoms, so that they cannot consist of unconnected parts. Note also how this represents the characterisation of appearances in space as *extensive magnitudes*. All spaces are given by first generating their parts, and subsequently representing these parts as unity trough the synthesising function of summation.

Now we would like to insist that in fact *our whole space* is connected, i.e., $\forall a, b \in A : C^*_A(a, b)$. However, taking space to be a path-connected structure of atomic regions can have some very undesirable consequences. Let $t = sum(A)$, suppose $|A| > 1$ and let $a \in A$ such that $A \setminus \{a\}$ is still path-connected. We can easily see that such an atom exists, for choose arbitrary $b \in A$ and let $l(c)$ for each $c \in A$ denote the length of the shortest path from b to c. Then we can choose any c for which $l(c)$ is maximal, and note that none of the shortest path from b to other nodes in A passes through c. Thus, $A \setminus \{c\}$ is path-connected and we can set $a = c$. Now, we would like to be able to define region $x_1 = sum(A \setminus \{a\})$. Note however that $C(x_1, a)$ so that $\forall y \in A : C(x_1, y)$ which implies $P(x_1, t)$ and thus $EQ(t, x_1)$ and even $P(a, x_1)$. This is very undesirable indeed. I thus add a 'sub-atomic' non-tangenial proper part inside every atomic region, to which no content or temporal structure is assigned and that is not intuited itself. The sub-atomic regions merely serve as representations of the 'inside', which must always exist since space is potentially infinitely divisible.

Axiom 3. *Given spatial structure* $\mathcal{S}$ *with atomic regions* A, *for every* $a \in A$ *there exists an* $a' \in SA$ *such that* $NTPP(a', a)$.

Recall that a non-tangenial proper part of a can only be connected to regions that overlap with a. Now we can properly distinguish larger regions so that the following holds:

Proposition 6. *Let* $a_1, a_2 \in S$, *then* $C^*(a_1, a_2)$, *where* C^* *transitively closes* C.

Proof. Since $u \in S$ and $\forall x \in S : C(x,u)$ we have $\forall x \in S : P(x,u)$, hence $\forall x : \neg PP(u,x)$. By axiom 2 then $u = sum(A_1)$ for $A_1 = \{b_1, \ldots, b_n\} \subseteq A$. Suppose $b \in A \setminus A_1$. Then consider the sub-atomic region c such that $NTTP(c,b)$. Since $C(c,u)$ we have by definition of the sum $\bigvee_{1 \leq i \leq n}(C(c,b_i))$. Consider the witness b_i for which $C(c,b_i)$. Since $P(c,b)$ we have $C(b,b_i)$, so that either $SO(b,b_i)$ or $EC(b,b_i)$. In the first case, axiom 1 implies $EQ(b,b_i)$. In the second case, it is prohibited that $EC(c,b_i)$ since $NTTP(c,b)$. But then since $C(c,b_i)$ we have $SO(c,b_i)$. Clearly however, $SO(c,b_i)$ implies $SO(b,b_i)$, which contradicts $EC(b,b_i)$. Thus, it must be the case that $EQ(b,b_i)$.

Since then for every $a, b \in A_1$ we have $C^*(a,b)$ by definition of the sum, and for every $c \in A$ we have $EQ(c,d)$ for some $d \in A_1$, the proposition follows. □

To ease computation, I further insist that for all $a \in A : PP(a',a) \wedge PP(a'',a) \rightarrow EQ(a',a'')$: every region has a single 'hypothetical inside', although it may have multiple names.

5.5 Dimension

Since Kant adhered to the general conception of space as three-dimensional, we must address the dimensionality of the RCC models constructed by the FAE. I have noted before that consistent sets of RCC-8 relations have models of connected regions in $\mathbb{R}^3$, but also in any higher dimension. For full RCC there is however no such guarantee. RCC allows for definition of unconnected regions and 'holes', conflicting with standard connected models of RCC-8 (Ligozat, 2013, p.322). A restriction of the spatial construction is thus needed. For reasons of computational efficiency I aim at representing two-dimensional space, but lay the conceptual groundwork for extension to higher dimensions.

From the viewpoint of topological models for RCC, we would like our definition of dimensionality to be a topological invariant (i.e., preserved under homeomorphisms) as well as weakly decreasing when taking the subspace. Furthermore, defining the boundary of a set x as $Cl(x) \setminus int(x)$, we would like $dim(boundary(x)) = dim(x) - 1$. Conversely, we might interpret spatial structures in terms of graphs. An overview of the correspondence between topologies and graphs in a spatial context can be found in the 'Handbook of Spatial Logic' (Aiello, Pratt-Hartmann, Van Benthem, et al., 2007). Both structures are subsumable under the larger class of *closure spaces*, and graph theory provides discrete counterparts for key topological notions. As a standard translation from topology τ on X to graph $G = (V,E)$ we can let $V = X$ and $E(x,y)$ iff $\exists A \in \tau : x \in A \wedge y \in A$. Alternatively one might apply the *specialization ordering* introduced in the previous chapter. In our point-free setting however, vertices represent regions (i.e., sets in the standard point-set interpretation). I define the connection graph induced by a spatial structure $\mathcal{S}$ as $G = (V,E)$, where $V = S$ and $E(a,b)$ iff $C(a,b)$. Since C is reflexive and symmetric, our graphs are *tolerance spaces* [1]. Let $N(x) = \{y \in V : R(x,y)\}$, and define the *punctured neighborhood* of a vertex x as $N_0(x) = N(x) \setminus \{x\}$. It is suggested by Michael Smyth and Julian Webster (Aiello et al., 2007, p.745) that at the level of *closure spaces*, the punctured neighborhood corresponds

[1] I use the term *tolerance space* for a reflexive and symmetric undirected graph, (Sossinsky, 1986)

to the topological boundary for open sets. They then provide the following definition of dimensionality:

Definition 23. *For tolerance space $G = (V, E)$, if G is empty, $dim(G) = -1$. Otherwise, $dim(G) = sup\{dim_G(x) : x \in V\}$, where for $x \in V$:*

$$dim_G(x) = \begin{cases} dim(N_0(x)) \text{ if } \exists y \in N_0(x) : N(x) \subseteq N(y) \\ dim(N_0(x)) + 1 \text{ otherwise} \end{cases}$$

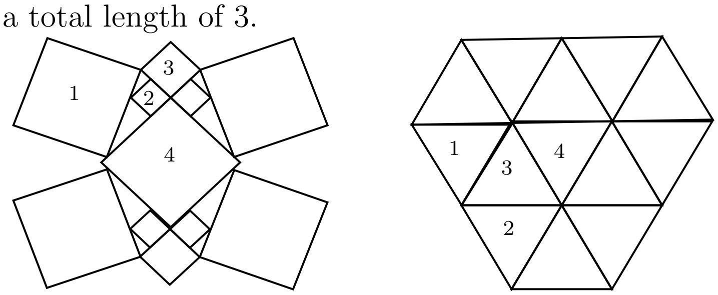
For example, a single clique (a set of vertices of which each pair is connected) has dimension 0. More generally, given a graph G we can consider the non-empty cliques that are intersections of maximal cliques under the name *cells*. Ordering these cliques by set-inclusion gives a partially ordered set $cell(G)$, on which we can define the *length* of an element as the length of the longest chain below it, and the *length* of a set as the maximal level over its elements, where the length of an empty set is -1. The following then holds:

Theorem 4. *Let G a non-empty tolerance space. Then $dim(G) = length(cell(G))$*

Proof. In (Aiello et al., 2007), p.746. □

Clearly, this measure is weakly decreasing under induced subgraphs as desired. Now we can apply this definition of dimensionality to the graphs induced by connected atomic regions in the two-dimensional plane. Since our regions are atomic, we do not allow overlaps. Still, without any restrictions on the size or shape of the regions we can obtain dimensions larger than 2, as is shown in figure 5.1. Both structures have dimension 3 from the perspective of *tolerance spaces*, while existing on the $2D$ plane. The left is made up of squares of different sizes, while the right is made up of triangles of the same shape and size.

Figure 5.1: Two figures of dimension 3. In both figures, $1, 2, 3, 4$ form a maximal clique of size 4 (length 0), $2, 3, 4$ form an intersection of two maximal cliques (length 1), $3, 4$ form an intersection of two such intersections (length 2) and 4 is the intersection of two intersections of length 2, giving a total length of 3.



Recalling that our atomic regions are to implement a magnitude following the schema of "addition of the homogeneous", I interpret them as regions of the same size and shape. The example above then shows that the shape in question matters. I choose to interpret atomic regions as squares, fix their orientation at the horizontal and vertical axis (i.e., I don't allow diamonds between the squares) and enforce that meeting squares have meeting corners (i.e. two squares with non-empty intersections have at least two corners that touch each other). The desired dimensionality of 2 then follows:

Theorem 5. *Let G be the connection graph induced by non-overlapping squares on the two-dimensional plane with the same size, the same directional orientation and where meeting squares have meeting corners. Then $dim(G) \leq 2$.*

Proof. Clearly, a clique of size 4 now only occurs if 4 squares meet with one corner in the same point, and a clique of size 5 cannot occur. Suppose $dim(G) > 2$. Then two cliques of size 4 must have an intersection of size 3, otherwise the length of $cell(G)$ can never be 3. Hence, 3 squares must meet in 2 distinct points with two distinct 4th squares. It is easy to see however that if 4 non-overlapping squares of equal size meet in one point, there can be no second point where 3 of these squares meet. The proof from geometric principles is left to the reader, but a simple picture (figure 5.2) suffices. □

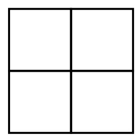
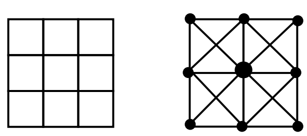


Figure 5.2: 4 non-overlapping squares of same size meeting in one point must form a partition of a single square; no 3 regions meet in a second point.

Moreover, if we draw a rectangular shape of at least 3×3 non-overlapping square atomic regions of equal size then we reach this upper bound: $dim(G) = 2$. Figure 5.3 shows such a shape of 3×3 and the induced connection graph, where vertices represent squares and edges represent the connection relation. Clearly, there are cliques of size 4 with intersections of size 2, all containing the same middle vertex. Thus $cell(G)$ has length 2 and $dim(G) = 2$ by theorem 4.

Figure 5.3: Left a 3 by 3 region of connected squares, right the induced connection graph G of dimension 2.



Since I have defined larger regions as sums over sets of atomic regions which constitute the magnitude of space, an analysis of the connection graphs induced by atomic regions is sufficient to evaluate the dimensionality of our spatial extension. In the next section I provide constraints to ensure that an interpretation of atomic regions in terms of non-overlapping squares is possible, thereby guaranteeing a dimension of at most 2.

5.6 Properties of connection graphs

In this section I define three graph-theoretic properties that provide a form of global consistency: the graph constructed by the FAE must be *locally sub-King*, *symmetry consistent* and must not contain any *inside corners*. The rest of this section then builds towards a proof that these properties ensure the constructed graph is a King-, Turband-King- or Torus-King graph, all of which have dimension ≤ 2. The reader who does not desire to plow through the proof in full may suffice by reading the definitions of the three conditions and graph-structures. The next chapter then continues by explaining how the three conditions are implemented to bring forth spatial experience.

Local orientation and dimension

I define two standard graphs as a basis for induced connection graphs. To make reasoning less convoluted, in what follows I analyse connection graphs only after removing all reflexive relations, hence $N_0(x) = N(x)$. I use $V(G), E(G)$ for edges and vertices of G. Since our constructed graph is induced by atomic regions, the axioms introduced above apply. Specifically, G must be path-connected as was shown in proposition 6. Now firstly, a *Grid* provides vertices with a two-dimensional interpretation, where only horizontally and vertically adjacent vertices are connected:

Definition 24 (Grid)**.** *Given integers p, q, **Grid graph** $H_{p,q}$ equals $P_p \times P_q$, where P_p is a path graph with p vertices labelled* 0 *to* $p-1$. *Label each vertex in $H_{p,q}$ with the corresponding element of* $\{0, \ldots, p-1\} \times \{0, \ldots, q-1\}$

Figure 5.3 shows that the connection graph induced by squares on a plane also contains 'diagonal' edges, so that we rather obtain the following connection graph:

Definition 25. *Given integers p, q, the **King graph** $K_{p,q}$ is a graph containing $H_{p,q}$ such that $V(K_{p,q}) = V(H_{p,q})$ and:*

$$E(K_{p,q}) = E(H_{p,q}), \cup, \{, ((i_x, i_y), (j_x, j_y)) : j_x = i_x + / -1 \ \wedge \ j_y = i_y + / -1 \wedge$$
$$0 \leq i_x, j_x \leq p-1 \ \wedge \ 0 \leq i_y, j_y \leq q-1\}$$

The graph in figure 5.3 is then a 3 by 3 king-graph. Since burdening the FAE with assigning each atomic region to a label (i_x, i_y) on the grid would be very inefficient, we enforce *locally* that each neighborhood $N(x)$ must be a subgraph of $K_{3,3}$. We would not insist that each neighborhood is equal to $K_{3,3}$, since this would prohibited the existence of edges (bounds). Naturally, while space as form of outer intuition allows for progressive infinity, space as concretely constructed in synthesis cannot be boundless, unless the ongoing process of successive synthesis is finished (see the first antinomy (A426/B454-A427/B455)).

Definition 26. *A graph G is **locally sub-King** if for each vertex $a \in V(G)$, there exists an injective assignment of $N(a)$ to a subset of the variables $x_0 \cdots x_7$, where the remaining variables are set to* $\mathbf{0}$*, such that:*

$$\forall 0 \leq i \leq 7 : x_i \neq \mathbf{0} \wedge x_{i+1} \neq \mathbf{0} \rightarrow C(x_i, x_{i+1})$$
$$\forall 0 \leq i \leq 7 : i \mod 2 = 0 \wedge x_i \neq \mathbf{0} \wedge x_{i+2} \neq \mathbf{0} \rightarrow C(x_i, x_{i+2})$$
$$\forall 0 \leq i, j \leq 7 : C(x_i, x_j) \rightarrow (x_i = x_{j+/-1} \vee (i \mod 2 = 0 \wedge x_i = x_{j+/-2}))$$

And, if for all $a \in V(G)$ the assignments to neighbours are surjective (i.e. $|N(a)| = 8$), the following further condition also holds for all $a \in V(G)$:

$$\forall\, 0 \leq i, j \leq 7 : i \mod 2 = 0 \rightarrow$$
$$((N(x_i) \cup N(x_{i+2}) \subseteq \{a, x_{i+1}\}) \wedge (N(x_i) \cup N(x_{i+4}) \subseteq \{a, N(x_{i+2}), N(x_{i+6})\}))$$

In the latter case, G is not only locally sub-King, but also ***locally King****.*

All indices are modulo 8. I denote variable x_i for the neighborhood of a with x_i^a. Note that this definition ensures $|N(x)| \leq 8$ by injectivity. Furthermore, the *locally sub-King* condition is sufficient to ensure two-dimensionality:

Theorem 6. *Let G a graph that is locally sub-King, then $dim(G) \leq 2$.*

Proof. Firstly note that the size of cliques in G is at most 4. To see this, suppose there exists a clique of size 5, then a vertex a has 4 pairwise connected neighbors. Neighbors with odd indices are connected to at most 2 other neighbors of a, so that each vertex in the clique must be assigned to an x_i^a with even i. But then x_0^a and x_4^a cannot be connected.

Hence, if $dim(G) = 3$ it must be the case by theorem 4 that two cliques of size 4 have an intersection of size 3. Thus, some vertex a must have two neighbors b, c that are connected to each other and are both connected to two distinct neighbors of a, d_1 and d_2. Again since b, c are connected to 3 neighbors of a they must be assigned to even x_i^a. Since b, c are connected the difference between the indices of the two associated variables is 2. But then the only other neighbor of a that can be connected to both is assigned to the odd index between the indices for b and c. Since the assignment to variables is injective: $d_1 = d_2$, a contradiction. Hence, $dim(G) \leq 2$. □

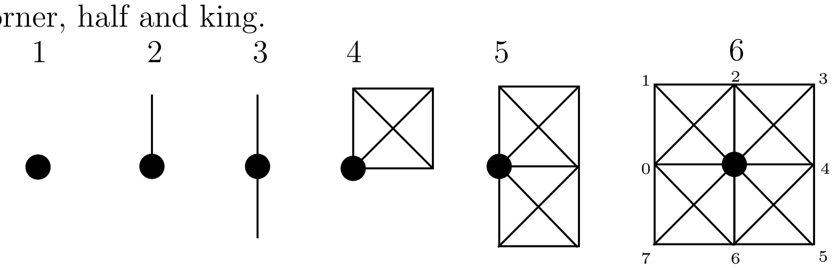
However, this condition is not yet sufficient for Kantian space. Specifically, *locally sub-king* graphs might represent a space that contains gaps. We hence further constrain neighborhoods so that they do not contain *inside corners*:

Definition 27. *A graph G that is locally sub-King* ***contains no inside corners*** *if for each vertex $a \in V(G)$ there exists an injective assignment of $N(a)$ to $x_1..x_7$ satisfying the* locally sub-king *condition so that the following holds:*

$$\forall 0 \leq i \leq 7 : i \mod 2 = 1 \wedge x_i \neq \mathbf{0} \rightarrow x_{i-1} \neq \mathbf{0} \wedge x_{i+1} \neq \mathbf{0} \tag{5.1}$$
$$\forall 0 \leq i \leq 7 : i \mod 2 = 0 \wedge x_i \neq \mathbf{0} \wedge x_{i+2} \neq \mathbf{0} \rightarrow x_{i+1} \neq \mathbf{0} \tag{5.2}$$

One may visualise this condition as only allowing one of the 6 neighborhoods below, although any mirroring or rotation by $k * 90°$ works.

Figure 5.4: Neighborhoods of locally sub-King graphs without inside corners. In turn: empty, single, line, corner, half and king.



Consistency of direction

While the conditions above ensure that each neighborhood is locally structured as a 'unit' of connected space, this is still not enough if we want to apply our intended directional interpretation. The indices for different neighborhoods might not at all match. For instance, we can have $x_1^a = b$ while $x_6^b = a$ so that a and b represent squares that are both connected by a single point (1 is odd) and by an edge (6 is even). The final condition I enforce on the constructed graph is thus that of symmetry in the neighborhood assignments:

Definition 28. $\forall 0 \leq i \leq 7$ *: let* $i^{-1} = i + 4$*. Then a locally sub-King graph without inside corners* G *is* ***symmetry consistent*** *if there exists an assignment of neighborhoods to* $x_0..x_7$ *satisfying definitions 26 and 27 so that* $\forall a \in V(G)\ \forall b \in N(a) : x_i^a = b \to x_{i^{-1}}^b = a$.

This is then sufficient to ensure at least locally an interpretation of the variables $x_0..x_7$ in terms of directions on a two-dimensional plane as in figure 5.4. The directed relations between neighbours of a vertex a match the directed relations of a to its neighbours. To see this, I define compositions of indices according to their directional interpretation:

Definition 29. *For* $0 \leq i \leq 7$ ***compositions*** *are as follows:*

$$i \mod 2 = 0 \to i \circ i + 2 = i + 1$$
$$i \mod 2 = 0 \to i \circ i - 2 = i - 1$$
$$i \mod 2 = 0 \to i \circ i + 3 = i + 2$$
$$i \mod 2 = 0 \to i \circ i - 3 = i - 2$$
$$i \mod 2 = 1 \to i \circ i + 3 = i + 1$$
$$i \mod 2 = 1 \to i \circ i - 3 = i - 1$$

Where for other pairs of directions the composition is undefined since it does not correspond with a connection on the 2d plane.

Note the agreement of these compositions with the numbering of neighbors in figure 5.4. We are now ready to prove the consistency of directions among neighbours. Unfortunately the proof is a rather tedious exercise in comparing indices. I refer to symmetry consistency (definition 28) as (SC) and to prohibition of inside corners (definition 27) as (¬IC):

Theorem 7. *Let* G *a graph that is locally sub-King and satisfies (SC) and (¬IC), and consider the assignment to indices satisfying all these properties. For all vertices* $a \in V(G)$*,* $b, c \in N(a)$ *and directions* i, j *where* $i \circ j$ *is defined:* $x_i^a = b \wedge x_j^b = c \to x_{i \circ j}^a = c$*. Furthermore, if* $i \circ j$ *is undefined:* $x_i^a = b \wedge x_j^b = c \to c \notin N(a)$*.*

Proof. Suppose $i \mod 2 = 0$. W.l.o.g. assume $i = 0$. Now suppose $j = i + / - 2$. Again w.l.o.g. let $j = 2$. By (SC) $x_{i^{-1}}^b = x_4^b = a$. Since $x_2^b = c$ then $C(a, c)$. By (¬IC) there must be a d such that $x_3^b = d$ and $C(a, d), C(c, d)$, i.e. a, b, c, d form a clique. Since $x_7^d = b$ (again by (SC)) we have either $x_6^d = a$ or $x_0^d = a$. In the first case, we have $x_2^a = d$ so that, since a, b, c, d is a clique, $x_1^a = x_{i \circ j}^a = c$ as desired. In the second case we have $x_4^a = d$ by (SC), contradicting $C(b, d)$.

Suppose instead that $i \mod 2 = 0$ and $j = i+/-3 \mod 8$. Again w.l.o.g. let $i = 0, j = 3$. By (SC) $x^b_4 = a$, and by (¬IC) and $x^b_3 = c$ there must be a d such that $x^b_2 = d$, so that again a, b, c, d form a clique. Since $x^c_7 = b$ (by (SC)) we have $x^c_6 = a$ or $x^c_0 = a$. In the first case, $x^a_2 = x^a_{i \circ j} = c$ by (SC) as desired. In the second case $x^a_4 = c$ which contradicts $C(a, b)$.

Now suppose $i \mod 2 = 1$ and $j = i + / - 3$. W.l.o.g. let $i = 1, j = 4$. By (SC), $x^b_5 = a$. By $(\neg IC)$ there must be a d such that $x^b_6 = d$ and again a, b, c, d form a clique. Since $C(b, c)$ we have $x^a_2 = c$ or $x^a_0 = c$. In the first case we are done since $1 \circ 4 = 2$. In the second case we have $x^c_4 = x^c_{0^{-1}} = a$ by (SC). Then by $x^a_1 = b$ and the previous case of compositionality proven above, it must be the case that $x^c_{4\circ 1} = x^c_2 = b$. However, since $x^b_4 = c$ we have $x^c_0 = b$, a contradiction.

Finally suppose $i \circ j$ is undefined. I show by cases that $c \notin N(a)$. Firstly suppose $i = j$. Then $x^b_{j^{-1}} = x^b_{j+4} = a$. But then a and c cannot be connected neighbours of b. Similarly, if $i = j + / - 1$ we have $x^b_{i^{-1}} = x^b_{j+/-3}$ so that a and c cannot be connected neighbours of b. If $i = j + / - 4$ we have $j = i^{-1}$. But then by injectivity of the assignment $c = a$ so that $c \notin N(a)$. The final remaining case is when $i \mod 2 = 1$ and $j = i+/-2$. Suppose w.l.o.g. that $i = 1$, $j = 3$. In this case by (SC) $x^b_5 = a$. By (¬ IC) there must exist a d such that $x^b_4 = d$. Note that $d \in N(a) \cap N(c)$. Now we can apply the proof of compositionality given above. Since $x^a_1 = b, x^b_4 = d$ we have $x^a_{1\circ 4} = x^a_2 = d$. Similarly, since by (SC) $x^d_0 = b$ and since $x^b_3 = c$ we have $x^d_{0\circ 3} = x^d_2 = c$. Since then $x^a_2 = d$ and $x^d_2 = c$ we have $c \notin N(a)$ as was shown above. □

Theorem 8. *Let G be locally sub-King so that (SC) and (¬IC) hold and consider the satisfying assignment of neighbors to indices. For all vertices $a \in V(G)$, $b, c \in N(a)$ and directions i, j where $i \circ j$ is defined: $x^a_i = b \wedge x^a_{i\circ j} = c \to x^b_j = c$.*

Proof. Immediate from theorem 1. Note that $i \circ j$ is $i + / - 1$ if i is odd and $i + / - 1$ or $i + / - 2$ if i is even. In both cases $C(b, c)$. Now suppose $x^b_l = c$ for some l such that $i \circ l$ is undefined, then by the second part of theorem 7 we have $c \notin N(a)$ which is a contradiction. Conversely suppose $x^b_l = c$ for some $l \neq j$ such that $i \circ l$ is defined. Then by the first part of theorem 7, $x^a_{i\circ l} = c$. Thus we have $c = x^a_{i\circ j} \neq x^a_{i\circ l} = c$, which is again a contradiction. □

Theorems 7 and 8 together show that each neighborhood can be drawn on a 3 × 3 grid so that the directional orientation of vertices corresponds to the intended interpretation of the indices.

Global properties

Now that I have made sure that the local conditions work as intended, I turn towards the global structure of the constructed graph. Ideally we do not merely want the assurance that each neighborhood has a local model with a directional interpretation, but we also want to know that our space as a whole can be interpreted as a two-dimensional spatial structure, preferably one without 'screws', 'wormholes' and other complex curvatures. In this section I show that this is indeed the case. Let's first define several canonical graphs, indices i_x, j_x are modulo p and indices i_y, j_y are modulo q:

Definition 30. *Given integers p, q, the **Turband graph** $Tu^{\delta}_{p,q}$ is a graph containing $H_{p,q}$ (the p by q grid) so that $V(T^{\delta}_{p,q}) = V(H_{p,q})$ and:*

$$E(Tu^{\delta}_{p,q}) = E(H_{p,q}) \cup \{((i_x, 0), (i_x + \delta, q - 1)) : 0 \leq i_x \leq p - 1\}$$

Definition 31. *Given integers p, q, the **Turband-King** graph $TuK^{\delta}_{p,q}$ is a graph containing $Tu^{\delta}_{p,q}$ such that $V(TuK^{\delta}_{p,q}) = V(Tu^{\delta}_{p,q}) = V(H_{p,q})$ and:*

$$E(TuK^{\delta}_{p,q}) = E(Tu^{\delta}_{p,q}) \cup E(K_{p,q}) \cup \{((i_x, 0), (i_x + \delta + / - 1, q - 1)) : 0 \leq i_x \leq p - 1\}$$

Intuitively, the Turband and Turband-King graphs are Grids or King-graphs of which the top and bottom vertices have been connected, possibly with a rotation of δ.

Definition 32. *Given integers p, q, the **Torus** graph $To^{\delta}_{p,q}$ is a graph containing $H_{p,q}$ such that $V(To^{\delta}_{p,q}) = V(H_{p,q})$ and:*

$$\begin{aligned} E(To^{\delta}_{p,q}) = E(H_{p,q}) &\cup \{((i_x, 0), (i_x + \delta, q - 1)) : 0 \leq i_x \leq p - 1\} \\ &\cup \{((0, i_y), (p - 1, i_y)) : 0 \leq i_y \leq q - 1\} \end{aligned}$$

Definition 33. *Given integers p, q, the **Torus-King** graph $ToK^{\delta}_{p,q}$ is a graph containing $To^{\delta}_{p,q}$ such that $V(ToK^{\delta}_{p,q}) = V(To^{\delta}_{p,q}) = V(H_{p,q})$ and:*

$$\begin{aligned} E(ToK^{\delta}_{p,q}) = E(To^{\delta}_{p,q}) \cup E(K_{p,q}) &\cup \{((i_x, 0), (i_x + \delta + / - 1, q - 1)) : 0 \leq i_x \leq p - 1\} \\ &\cup \{((0, i_y), (p - 1, i_y + / - 1)) : 0 \leq i_y \leq q - 1\} \end{aligned}$$

The torus is a graph-representation of a doughnut. It is a turband where also the left and right boundaries have been connected, but without any rotation δ_2 on this axis. The turband and torus are examples of connection graphs induced by two-dimensional planes that are folded without any 'screws'. An example of a graph where this is not the case is the Klein bottle:

Definition 34. *Given integers p, q and $i \in \{0, 1, 2\}$, where $i = 1 \leftrightarrow p \mod 2 = 1$, the **Klein bottle** $KB^{i}_{p,q}$ is a graph containing $H_{p,q}$ such that $V(KB^{i}_{p,q}) = V(H_{p,q})$, and:*
if $i \in \{0, 1\}$:

$$E(KB^{i}_{p,q}) = E(Tu^{0}_{p,q}) \cup \{((i_x, 0), (p - i_x - 1, q - 1)) : 0 \leq i_x \leq p - 1\}$$

While if $i = 2$:

$$E(KB^{i}_{p,q}) = E(Tu^{0}_{p,q}) \cup \{((i_x, 0), (p - i_x, q - 1)) : 0 \leq i_x \leq p - 1\}$$

The torus and Klein bottle graphs are embeddable in the associated topological structures shown in figure 5.5

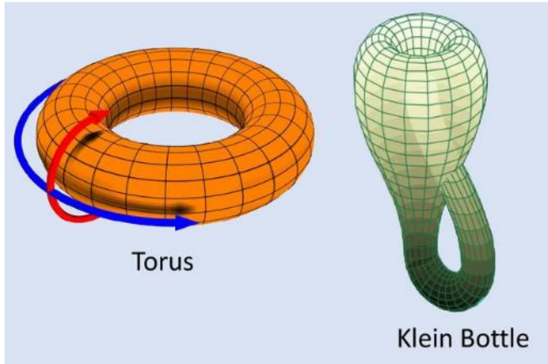


Figure 5.5: Topological torus and Klein bottle

In what follows, I show that the graphs constructed by the FAE are isomorphic to either a King-, Turband-King or Torus-King graph. Screws like that of the Klein-bottle are not allowed. This requires a rather long sequence of proofs. In these proofs I always take G to be finite, locally-sub-King, symmetry-consistent, path-connected and without inside corners, and I analyse the assignment of neighborhoods to indices that satisfies these properties.

I associate each variable index i in a neighborhood with a change in x, y coordinates $\delta_{xy}(i)$, in the intuitive manner that matches figure 5.4. Thus $\delta_{xy}(0) = (-1, 0), \delta_{xy}(1) = (-1, 1)$ etc.. I write $dir(e)$, for the index associated with edges, i.e., $dir((a, b)) = i$ if $x_i^a = b$. Define a path p as a finite sequence of consecutive vertices and directed edges $(v_1 e_1 v_2 \ldots, e_{n-1} v_n)$, such that e_i is an edge from v_i to v_{i+1} for $1 \leq i < n$ and let $\delta_{xy}(p) = \sum_{e \in p \cap E(G)} \delta_{xy}(dir(e))$. I first prove that all paths have counterparts with the same traversal and without diagonals:

Lemma 7. *For each path $p = (v_1 \ldots v_n)$ over edges of G, there exists a path p' from v_1 to v_n such that $\delta_{xy}(p) = \delta_{xy}(p')$ and all edges in p' have direction $0, 2, 4$ or 6.*

Proof. This is clear from the exclusion of inside corners. Let e_i be the first edge in p such that $dir(e_i) \notin \{0, 2, 4, 6\}$. By ($\neg$IC), there also exists a vertex v_j such that $x^{v_i}_{(dir(e_i)+1)} = v_j$. By the locally sub-King condition $C(vj, v_{i+1})$. But then note that $dir(e_i) = dir(e_i) + 1 \circ dir(e_i) - 1$ (since for even directions d we have $d \circ d - 2 = d - 1$). Thus, by theorem 8 there exists a path of length two from v_i to v_j to v_{i+1} of which the two edges have directions $dir(e_i) + 1$ and $dir(e_i) - 1$. Name these edges e', e''. It is then clear from the x, y changes associated with the indices that $\delta_{xy}(e_i) = \delta_{xy}(v_i e' v_j e'' v_{i+1})$. If we substitute the latter path for the sub-path $(v_i e_i v_{i+1})$ in p, we obtain a path that has one less edge with direction $1, 3, 5$ or 7 and the same $\delta_{x,y}$. By iterating this procedure we obtain p'. □

Now we can show that paths with $(0, 0)$ traversal are reflexive:

Lemma 8. *For each path $p = (v_1 \ldots v_n)$ over edges of G, if $\delta_{xy}(p) = (0, 0)$ then $v_1 = v_n$.*

Proof. I assume the negation and derive a contradiction. Suppose a path $p = (v_1 \ldots v_n)$ exists such that $\delta_{x,y}(p) = (0, 0)$ but $v_1 \neq v_n$. By the foregoing lemma there exists a path $p' = (v_1 \ldots v_n)$ such that $\delta_{x,y}(p') = (0, 0)$ and for all edges e in p': $dir(e) \in \{0, 2, 4, 6\}$. Let p' be the shortest such path (if there are several shortest path of equal length, let p' be an arbitrary one of them). We can assume p' has no cycles, i.e. there are no strict sub-paths $q = (w_1 \dots w_m)$ of p' such that $m < n$ and $\delta_{xy}(q) = (0,0)$, for if $w_1 \neq w_m$ we can let $p' = q$ and continue towards a contradiction, and if $w_1 = w_m$ we can remove $(w_1 \dots w_{m-1} e_{m-1})$ from p' to obtain a shorter path with the same starting and end vertex of p and the same $\delta_{x,y}$. Furthermore, we can assume p' is not a 'line', i.e. a path where all edges e have direction $dir(e) \in \{0,4\}$ or all edges have direction $dir(e) \in \{2,6\}$. For suppose that we have $e_i, e_{i+1} \in p$ such that $dir(e_{i+1}) = dir(e_i) + / - 4$. Then clearly $q = (v_i e_i v_{i+1} e_{i+1} v_{i+2})$ is a cycle with $\delta_{x,y}(q) = (0,0)$. Thus, q cannot be a strict sub path of p' which implies $q = p'$. But then since e_i and e_{i+1} have opposite directions it follows from constraint (SC) that $v_i = v_{i+2}$, contradicting that the start and end of p' are distinct.

Hence, if we draw p' on the x, y plane we obtain the boundary of a single connected two-dimensional region. Note that the start and end of this boundary on the x, y plane are the same, and that no edges cross each other, for this would imply the existence of a cycle in p'. Denote the surface of the region enclosed by this boundary by m. I now show that from p' we can iteratively construct paths $q_1, \cdots q_k$ where $k \leq m$ such that the following holds. Here, $l(p)$ denotes the number of edges in p, $s(p)$ denotes the size of the surface enclosed by p if it is indeed the boundary of a two-dimensional connected region, and I let $p' = q_0$ for convenience:

$$\begin{aligned} \forall 1 \leq i \leq m : q_i = (v_1 \dots v_n) \\ \wedge\, \delta_{x,y}(q_i) = (0,0). \\ \wedge\, l(q_i) \leq l(q_{i-1}) \\ \wedge\, s(q_i) < s(q_{i-1}) \end{aligned}$$

In words, the size of the surface enclosed by q_i goes to 0 as the sequence progresses. Note that the surface enclosed by p' has rectangular corners, so that it must have at least 4 'convex corners' i.e., corners of which the enclosed surface touches the 90 degree corner and not the 270 degree corner (see figure 5.6). At least 3 of these convex corners must be between two successive edges in p', since at most 1 of them is between e_1 and e_{n-1}. Take one of these corners and let e_i, e_{i+1} be the associated edges. Clearly, $dir(e_i) = dir(e_{i+1}) + / - 2$. Since all edges in p' have an even direction $dir(e_i) \circ dir(e_{i+1})$ is defined, so that by theorem 7 there also exists an edge e_j from v_i to v_{i+2} with $dir(e_j) = dir(e_i) \circ dir(e_{i+1})$. Suppose w.l.o.g. that $dir(e_i) = 0, dir(e_{i+1}) = 2$. Then $dir(e_j) = 1$. By ($\neg$IC) there also exists a vertex w such that v_i is connected to w by an edge e_w with direction 2. Then since $1 = 2 \circ 0$ there must by theorem 8 also exists an edge $e_{w'}$ from w to v_{i+2} such that $dir(e_{w'}) = 0$.

Hence, we can 'cut' the corner by replacing the sequence $(v_i\, e_i\, v_{i+1}\, e_{i+1}\, v_{i+2})$ by the sequence $(v_i\, e_w\, w\, e_{w'}\, v_{i+2})$ in p'. The resulting path q_1 clearly has the same $\delta_{x,y}$ as p' and has the same length. Furthermore, since the original 'cut' corner was convex and our new path now goes through the other shared neighbour of v_i and v_{i+2} we have subtracted 1 from the enclosed surface m. Since the q_1 now runs through w instead of v_{i+1} one might ask whether it can contain a cyclic sub-path r with $\delta_{x,y}(r) = (0,0)$. However, as argued before, the existence of such a cycle would imply that q_1 is not a shortest path with distinct start and end vertices such that $\delta_{x,y}(p') = (0,0)$. Either r is a shorter path with a distinct start and end, or we can cut r from q_1.

We iteratively continue this process, constructing paths with the same start and end as p' (or alternatively starting with a new shorter p'), the same $\delta_{x,y} = (0,0)$, and enclosing smaller and smaller surfaces. Now finally we reach a path q_j with $j \leq m$ that encloses a surface of size 0. Hence, q_j must be a line from v_1 to v_n where $v_1 \neq v_n$ and $\delta_{x,y}(q_j) = (0,0)$. However, as I have shown above, such a path cannot exist. We thus reach a contradiction, proving the lemma. □

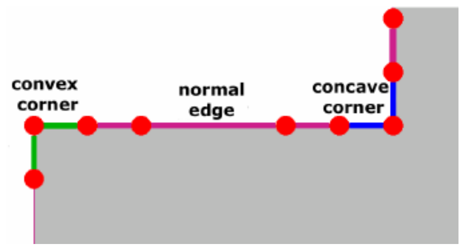


Figure 5.6: Concave and convex corners, (Gao, Gu, & Zakhor, 2008)

Using this lemma we can see that if there exists a path from vertex a to itself, then every path with the same traversal is reflexive.

Lemma 9. *If for some vertex $a \in V(G)$, there exists a path $p = (a \dots a)$ with $\delta_{x,y}(p) = (dx, dy)$, then for every path $q = (b \dots c)$ with $\delta_{xy}(q) = (dx, dy)$, we have $b = c$.*

Proof. Let the inverse of a path p from x to y denote the path from y to x traversing the edges of p in reverse order. Since G is path-connected, there must exist a path r from b to a. Thus, there exists a path $s = (rp^{-1}r^{-1})$ from b to b with $\delta_{x,y}(s) = (-dx, -dy)$. Thus, there exists a path $t = (sq)$ from b to c with $\delta_{x,y}(t) = (0,0)$. By Lemma 8 then $b = c$. □

From locally King to Torus

We are now ready for our first result on the global structure of G.

Theorem 9. *If G is a locally King-graph of size $|V(g)| = n$, then it is isomorphic to a Torus-King graph $ToK^{\delta}_{p,q}$ where $p \cdot q = n$.*

The proof makes use of an analysis of locally Grid-graphs made by Marquez et. al (2003). Locally Grid-graphs are graphs where for each vertex v there exists an injective assignment from $N(x)$ to $x_1, \dots x_4$ and we have four different vertices $y_1 \dots y_4$ such that:

$$N(x_i) \cap N(x_{i+1}) = \{v, y_i\}$$
$$N(x_i) \cup N(x_{i+2}) = \{x\}$$

Clearly, if G is a locally King graph and we remove all edges of odd direction the result is a locally Grid graph of which all edges have even labels, henceforth denoted as G'. Marques et.al further make use of the concepts of opposite walks and opposite cycles:

Definition 35. *An opposite walk is a path* $(v_1 e_1 \dots v_n)$ *such that* e_i *is* opposite *to* e_{i+1} *for* $1 \leq i < n$*, meaning that there is no square of four edges containing both* e_i *and* e_{i+1}*, and such that* v_n *is the first vertex equal to an earlier vertex in the path.*

Definition 36. *An opposite cycle is a path* $(v_1 e_1 \dots v_1)$ *such that* e_i *is* opposite *to* e_{i+1} *for* $1 \leq i < n$.

Lemma 10. *A locally Grid graph* G *for which all opposite walks are opposite cycles is isomorphic to either:*

1. *Torus* $To^{\delta}_{p,q}$ *with* $p \geq 5, q \geq 1$*, where* $\delta \geq 4$ *if* $q = 1$*, and* $\delta + q \geq 6$ *if* $q \in \{2, 3\}$.
2. *Klein bottle* $KB^i_{p,q}$ *with* $p \geq 5$*,* $i = p \mod 2$ *and* $q \geq 4 + \lceil i/2 \rceil$.

In both cases, $p \cdot q = |V(G)|$.

Proof. In Marquez et.al. Note that we can assume $\delta \leq p/2$ by symmetry. □

Now it follows from Lemma 9 that for our locally Grid graph G' all opposite walks are opposite cycles.

Proof. Note that in our directed setting an opposite walk equals a walk along edges with the same direction. If $dir(e_{i+1}) = dir(e_i) + / - 2$ for some edge e_i in opposite walk p, then by ($\neg$ IC) e_i and e_{i+1} are part of the same square. If $dir(e_{i+1}) = dir(e_i) + / - 4$, then by (SC) we have $v_i = v_{i+2}$ so that e_i and e_{i+1} are also part of the same square, contradicting the definition of *opposite* edges.

Whith this in mind, let $p = (v_1 \dots v_n)$ be any opposite walk. Now $\delta_{xy}(p) = (n-1) * \delta_{x,y}(e)$ for any edge $e \in p$. By definition of opposite walks: $v_n = v_i$ for some $1 \leq i < n$. Thus, there exists a path q from v_n to v_n such that $\delta_{xy}(q) = (n - i) * (\delta_{x,y}(e))$. By lemma 9, every path with this δ_{xy} starts and ends in the same vertex. Thus, since the initial segment $r = (v_1 \cdots v_{n-i+1})$ of p is such a path, we have $v_1 = v_{n-i+1}$. Since v_n is the first vertex in the walk that is equal to an earlier vertex, by definition of opposite walks, it follows that $i = 1$. Thus, p is an opposite cycle. □

Applying Lemma 10 then G' is a Torus or Klein bottle. Since definitions 32 or 34 thus apply, $V(G') = H_{p,q}$ for some p, q. It remains to be shown that G' must in fact be a Torus:

Firstly, I prove that the directions on the p, q grid must be isomorphic to the directional indices $0, 2, 4, 6$ of the edges of G', in the sense that each of the four basis directions $\delta \in \{(0, 1), (0, -1), (1, 0), (-1, 0)\}$ maps to one of the 4 indices by injective function f so that $b - a = \delta$ implies $x^a_{f(\delta)} = b$

Proof. To see this, consider vertices $(0, 0)$ and $(1, 0)$, and suppose w.l.o.g. that the edge from $(0, 0)$ to $(1, 0)$ is labelled 4. By (SC) there must be an edge from $(1, 0)$ to $(0, 0)$ labelled 0. Note then that $(0, 0)$ and $(2, 0)$ do not share a neighbour apart from $(1, 0)$ since $p \geq 5$, if G' is a Torus and $q = 1$ then $\delta \geq 4$ (note that $p \geq 9$ in this case), and if G' is a Klein bottle then $q \geq 4$. Now if the edge $((1, 0), (2, 0))$ would have an index other than 4, ($\neg$ IC) along with compositionality (theorem 7) would require that $(0,0)$ and $(2,0)$ share at least two neighbors. Thus $((1,0),(2,0))$ is also labeled 4. We can repeat this argument to show that all edges $((i,0),(i+1,0))$ are labelled 4 and all edges $((i,0),(i-1,0)$ are labelled 0. Now if $q=1$ then we are done, all edged associated with a δ on the grid have been assigned a suitable index.

Suppose then $q>1$, it must be the case that $(1,0)$ and $(0,1)$ have two common neighbors $(0,0)$ and $(1,1)$. By the last constraint of the locally sub-King condition, this implies that since $((0,0),(1,0))$ has label 4 it must be the case that $((0,0),(1,0))$ has label $l \in \{2,6\}$. Suppose w.l.o.g. that $l=2$. Since $(0,0)$ and $(1,1)$ share two common neighbors, it follows then by compositionality that $((0,1),(1,1))$ has label $l' \in \{0,4\}$. Suppose $l'=0$, then by compositionality there exists edge $((0,0),(1,1))$ with label $2 \circ 0 = 1$ in our original graph G. But then $((0,1),(1,1))$ has label l'' such that $4 \circ l'' = 1$, and such an l'' does not exist. Thus $l'=4$. Using a similar argument by diagonals in G, we can see $((0,1),(1,1))$ has label 2. But then we can iterate this argument for edges $((2,0),(2,1))$ and $((1,1),(2,1))$. This shows $((i,0),(i,1))$ has label 2 for $0 \le i \le p$ and $((i,1),(i+1,1)$ has label 4 for $0 \le i < p$, with the opposite δ directions labeled 6 and 0 respectively by (SC).

Suppose $q=3$, then again since $(0,2)$ and $(1,1)$ share two neighbors $(0,1)$ and $(1,2)$, and since $((0,1),(1,1)$ has label 4, $((0,1),(0,2)$ has label $l \in \{2,6\}$. $l=6$ gives $(0,0)=(0,2)$ contradicting $q=2$. Thus, $l=2$. Now we can apply the exact same argument as for the row below to find $((i,1),(i,2))$ has label 2 for $0 \le i \le p$ and $((i,2),(i+1,2)$ has label 4 for $0 \le i < p$, with the opposite δ directions labeled 6 and 0 respectively by (SC). Iterating this process per row shows that every edge $(i,j),(i+1,j)$ has label 4, $(i,j),(i-1,j)$ has label 0, $(i,j),(i,j+1)$ has label 2 and $(i,j),(i,j-1)$ has label 6. Note that we could have assigned other labels to the δ, as long as the inverse operation is preserved. □

Now I can show $G' \neq KB^i_{p,q}$ for $p \ge 5$, $i = p \mod 2$ and $q \ge 4 + \lceil i/2 \rceil$

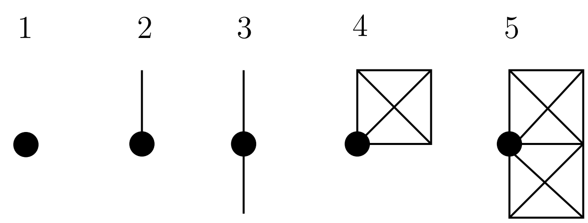
Proof. Suppose $G' = KB^i_{p,q}$ with the parameters as specified above. Consider vertex $(1,0)$. By the above it is connected to $(0,0)$ and $(2,0)$ by edges labelled l and l^{-1} (for $l \in \{0,2,4,6\}$), i.e., $f((-1,0)) = l^{-1}, f((1,0)) = l$. Since $q \ge 4$, $(1,0)$ is also connected to $(1,1)$ by an edge labeled r, distinct from l and l^{-1}. Now if $i \in \{0,1\}$, then $(1,0)$ is connected to $(p-2,q-1)$ by an edge labeled r^{-1} (the only remaining index), while $(2,0)$ is connected to $(p-3,q-1)$ by an edge with the same label r^{-1}. Note further that $x_l^{(p-3,q-1)} = (p-2,q-1)$ by the isomorphism of directions on the grid to indices proven above. Hence, in our original graph G, by compositionality (theorem 7) there exists an edge $((1,0),(p-3,q-1))$ with label $r^{-1} \circ l^{-1}$ and an edge with label $l \circ r^{-1}$. Since these are two distinct labels this contradicts that G is a locally King graph. If $i=2$ we can construct a similar argument. $(1,0)$ must then be connected to $(p-2,q-1)$ by two edges with distinct labels. □

This proves that if G is a locally King graph, then G' is a Torus graph. But then it is easy to see by compositionality that G is a Torus-King graph. The only edges we have to add to G' to produce G are those between vertices a,c where for some b and even indices i,j: $x_i^a = b, x_j^b = c$ and $i \neq j, i \neq j^{-1}$. Using the locally King-condition, these are then only the vertices such that $|N(a) \cup N(c)| = 2$ in G'. We must thus only add the edges $((i_x,i_y),(j_x,j_y))$

such that $j_x = i_x + / - 1, j_y = i_y + / - 1$ and $0 \leq i_x, j_x \leq p - 1, 0 \leq i_y, j_y \leq q - 1$, the edges $((i_x, 0), (i_x + \delta + / - 1, q - 1))$ where $0 \leq i_x \leq p - 1$, and the edges $((0, i_y), (p - 1, i_y + / - 1))$ where $0 \leq i_y \leq q - 1$. Clearly this matches definition 33. This concludes the proof of theorem 9.

From strictly locally sub-King to King graph or Turband

Now I continue by showing that if G is not locally King, it must be a King-graph or Turband-King graph. In this case, there must be at least one vertex v of which the neighborhood is not a 3×3 King graph. By constraint (¬ IC), this vertex can have one of 5 possible neighborhood structures:



I proceed by cases over the neighborhood of v:

If v has neighborhood **1**, path-connectedness ensures that our graph is a single vertex and hence a 1 by 1 King-graph.

If v has neighborhood **2**, it is connected to a single node w by an edge with label l. Note then that by (¬IC), w must have neighborhood 2 or 3, as is the case for any vertex connected to w. By simple induction, every vertex in G then has neighborhood structure 2 or 3. Moreover, there exists a path from v through all vertices in G such that each edge has label l (all other labels are excluded by compositionality, (¬IC) and (SC)). Suppose that a vertex s distinct from v has neighborhood 2. Then there must be path from v to s of length $x \geq 1$, such that all intermediate vertices have neighborhood 3. We thus have a King graph $K_{0,x}$. Note that it cannot be the case that more than two vertices have neighborhood 2.

Suppose conversely that only v has neighborhood 2, while all other vertices have neighborhood 3. Then we have an infinite path from v in which no vertex occurs twice, i.e., there are no cycles. This follows from lemma 9: if there exists a cycle in the path from v over l-labelled edges of length x, then v must be equal to the xth vertex in this path, which is impossible since v has neighborhood 2 while x has neighborhood 3. But then since G is finite such an infinite cycle-free path cannot exist. Thus there must exist another vertex with neighborhood 2.

If v has neighborhood **3**, again all other vertices must have neighborhood 2 or 3. If one vertex has neighborhood 2 we are in the same case as before. However, all vertices might have neighborhood 3, forming a cycle over edges with the same label l. There then exists a path from v to v over edges labelled l that visits all vertices in G since G is connected. After at most $V(G)$ edges the path must reach a vertex for the second time. But then note that by lemma 9, every path in the same direction of the same length must start and end in the same vertex. Then it must be the case that after exactly $V(G)$ edges labelled l the same vertex is reached. We can thus place our vertices on a grid where $p = 1, q = |V(G)|$ so that $(0, V(G) - 1)$ is connected to $(0, 0)$. G is then a Turband-King graph $Tu^0_{1,|V(G)|}$.

If v has neighborhood **4** we can apply a similar argument as when v has neighborhood 2, albeit a bit more complicated. We construct a Turband-King graph by drawing it directly on a p by q grid. We place v on $(0, 0)$. The vertices connected to v by edges with even labels must themselves have neighborhood 4 or 5 by $(\neg IC)$. Take one of these vertices w and let l the label of the edge from v to w. We place w on $(1, 0)$. Using a similar argument as when v has neighborhood 2, we can construct a path along the x axis of ≥ 1 edges labelled l. Note that this path again cannot contain cycles, since by 9 the existence of a subcycle would imply that every subpath of the same length would be a cycle, while v cannot occur twice in this path since it has no outgoing edge with direction l^{-1}. This then also implies that our path is finite and ends in a corner with neighborhood 4. Using the same argument we can construct a finite path from v in the other 'even' direction (name the associated label r) along the y axis. Note that no pair of vertices in the two paths have the same neighborhood: vertices in the first path have edges labelled l, l^{-1}, while vertices in the second path have edges labelled r, r^{-1}, so that all vertices in the two paths must be distinct (including the three corners with edge indices r and l, l^{-1} and r, and r^{-1} and l respectively). But then, compositionality requires us to fill in the whole p, q grid, where each edge is labelled as is required by compositionality. Note lemma 8 ensures that while filling in G on this rectangular Grid using the directions of the labelled edges we do not place two distinct vertices on the same coordinate, since this would imply the existence of a path p between these two vertices with $\delta_{x,y}(p) = (0, 0)$. Also note that the drawn graph must in fact be of rectangular shape, since from the two corners at the end of our constructed paths we can again follow two more finite 'edge paths'. Clearly the resulting rectangular graph must include all vertices, since G is connected. Finally, it cannot be the case that one vertex s is now placed on the two distinct coordinates $(x_1, y_1), (x_2, y_2)$ for there would then be a path from s to s with $\delta_{x,y}(p) \neq (0, 0)$, (i.e., $\delta_{x,y}(p) = (x_2 - x_1, y_2 - y_1)$ if $l = 4, r = 0$ but rotations are possible). But then we can iterate this path $|V(G) + 1|$ times to construct path $p^{|V(G)|+1}$ of which the absolute traversal in either the x or y direction must be larger than the size of G, i.e. s is beyond one of the four borders of G which is impossible. Thus in this case G must be a King graph.

Finally, suppose v has neighborhood **5**. This case is analogous to the easier case where v has neighborhood 3. There can be only one pair of mutually inverse edge-labels that are used in the neighborhood of v, name these l and l^{-1}. Note again that every vertex reached from v by edges labelled l must have neighborhood 4 or 5, while the same holds for l^{-1}. If G has a vertex with neighborhood 4 we are in the previous case so that G is a King graph. Otherwise, there must in fact be an infinite path in both directions. Note again that by lemma 9 there exists some cycle-length c such that each path of c edges in direction l (or l^{-1}) is a cycle. We place w at coordinate $(0, 0)$ and draw the y axis by following the path from w in the l direction for $c - 1$ edges. Clearly $(0, 0)$ and $(0, c - 1)$ are connected. Now let r be the other even direction in the neighborhood of v (i.e. $l + / - 2$). Note that each vertex reached following edges labelled r from v must have neighborhood 5 or a 3 by 3 King graph (numbered 6 in 5.4). Since no vertex in this direction has the same neighborhood as v there cannot be any cycles in this direction. Hence, there exists a finite path to 'the other edge'. We draw this path with length d along the x axis. Now we can again fill our rectangular grid as is required by compositionality. ($\neg$IC) ensures that every path of length d from a vertex on the y axis in the r direction reaches another 'edge vertex'. Lemma 8 again ensures that we do not fill in 2 vertices at the same coordinate. Furthermore, I have already argued that there can be no cycle by a path along the y axis of length $< c$. Hence, if compositionality would require us to place one vertex s at two coordinates there would be a non-zero distance along the x axis between these two coordinates. Following this path more than d times would then imply that s is 'beyond' the border at $x = d$, which is impossible due to prohibition of inside corners.

It remains to be shown that in the last case the bottom and top row of our grid are connected in a manner that fits the definition of a King turband Graph. Note that $(0, c-1)$ is connected to $(0, 0)$ by an $l-$edge. Since every path of c many l edges is a cycle (by lemma 9) this is the case for every $0 \leq x \leq r$. The associated diagonal edges follow by compositionality. G is thus a Turband King graph with $\delta = 0$.

Conclusion and Evaluation

This finishes the proof of the main theorem:

Theorem 10. *Let G a finite path-connected irreflexive locally sub-King tolerance graph that is symmetry consistent and contains no inside corners. Then G is isomorphic to a King-graph, Turband-King graph or Torus-King graph.*

We may now evaluate to what extent these spatial structures correspond to Kantian space. Of course, space for Kant is three-dimensional. In the current implementation construction of 3D graphs is however computationally very demanding, although the approach introduced here may from a theoretical standpoint very well be extended to three-dimensional structures. Interpreting the FAE then as a computational agent that constructs two-dimensional space it seems fitting that it may find loops such as that of the turband in its spacial structure. When moving round the surface of a house in a single direction for instance, one may at some point find herself back at the place where she started. A downside of allowing torusses and turbands is that the constructed space is bounded: no new vertices can be added to a torus without changing it to a non-torus. We thus see that if an artificial agent builds spatial structures of this form, it is not at all guaranteed that new sensations can be incorporated without changing the existing spatial information. If the agent wants to extend a spatial torus, for instance by making it wider, it must rebuild all edges along the bounds of the grid. This problem is then in a way analogous to what was said in the previous chapter about embeddings: there exists a tension between global consistency and successive synthesis. Again the issue is not critical for the FAE because it does not have a successive synthesis. Now for time the tension was limited in the sense that only the representation of time as a whole must be rebuilt if the scope of synthesis becomes wider. For space we see however that more intensive adjustments may be necessary: if a region is added it must be joined by neighboring regions to prevent gaps, and if cycles in space are elongated at a certain point this requires similar elongation throughout the whole of space to maintain consistency. On the other hand, one may if this is desirable easily restrict the currently developed spatial construction to two-dimensional *map-graphs* (Chen, Grigni, & Papadimitriou, 2002), by simply insisting that at least one vertex has neighborhood 4 from figure 5.4. From a topological point of view one might finally insist that constructed space should not merely be *connected*, but also *simply connected*: no holes pass all the way through it, such as the whole in the center of a torus. However, in the interpretation of an agent exploring the surface of a house, it seems sensible that one encounters impenetrable barriers around which space wraps itself.

This then concludes the theoretical analysis of space. In the next chapter this spatial structure is implemented in the Figurative Apperception Engine.

Chapter 6

Space: Implementation

Without further ado, I turn towards the construction of spatial structures implemented in the FAE. Importantly, the process operates in parallel with the temporal construction expounded earlier. The same events that are provided with temporal structure are also related by means of the connection relation C. Furthermore, the same embeddings of the manifolds distinguished by moments into larger manifolds under the amalgamation condition is still in effect. The important addition is that the input events are now assigned to *atomic regions*, and manifolds are spatially represented as sums over the *atomic regions* they contain. The spatial orientation among manifolds is constructed by means of *motion* in accordance with what was said in the introduction: Kant associates synthesis in space with motion as "action of the subject" (B155). In the example behavior at the end of this chapter, I show how space can be applied to resolve the extent to which objects must be given as input to the AE, alternatively constructing objects as covers over spatially coherent intuitions. I also combine the spatial and temporal constructions in application to Kant's example of the freezing of water.

6.1 Content in space

I now include a third type of content to CONTENTLOG: $(p,(v1,v2))$ where p is a predicate and v_1, v_2 are numerical values. Events with this content do not refer to objects, but represent mere *intensive magnitudes* in the sense of Kant's Anticipations of Perception. The aim of the FAE will then be to construct objects on the conceptual level that explain the input manifold under constraints of spatial coherence. While intuition is then that through which an object is *given* (instead of merely thought), the *object of experience* only comes into existence as the result of a unifying synthesis. This partly addresses the limitations of the AE regarding objects that were explained in chapter 2. I name the extended language CONTENTLOG^s, but do not repeat the full definition here.

6.2 Spatial structures

Implementation of the RCC is done trough straightforward translation into ASP constraints. Input events are now also called *atomic spaces*, and provided with *sub-atomic spaces* as insides (represented by the negative of the event number). Both types of spaces are subsumed under the more general term *element*. Now the C relation is guessed between elements, under constraint of axioms 1 and 3 (atomic spaces do not overlap, and have one non-tangenial proper part):

```
atomic_space(E) :-
    sense_input_event(E,_,_).

sub_atomic_space(-E) :-
    atomic_space(E).

has_event(M,-E) :-
    has_event(M,E), atomic_space(E).

prop_part(-E, E) :-
    atomic_space(E).

:-  sub_atomic_space(A), connected(E, A),
    not sub_atomic_space(E), not part(-A,E).

space_equal(E1,E2) :-
    atomic_space(E1), atomic_space(E2), space_overlap(E1,E2).

space_equal(A1,A2) :-
    sub_atomic_space(A1), sub_atomic_space(A2), connected(A1,A2).

element(E) :- atomic_space(E).
element(E) :- sub_atomic_space(E).

{connected(E1,E2)} :- element(E1), element(E2).
```

Program 6.1: Construction of atomic and sub-atomic spaces and the associated connection relation

I assign an index $0 \leq i \leq 7$ to each connection between atomic regions and ensure the locally sub-king condition as well as symmetry consistency and prohibition of inside corners. The program implementing this is rather technical and has been included in the appendix. Now to each manifold I assign a 'space_cover_event' that is the sum over its atomic events:

```
space_cover_event((T,0),E+T) :-
    has_event((T,0),_), time_cover_event(E).

% Sum definition: connection implies connection with part
:- has_prop_part(E1), has_prop_part(E2),
    not atomic_space(E1), not atomic_space(E2),
    connected(E1, E2), not connected(P1,P2) :
    prop_part(P1,E1), prop_part(P2,E2).

% Sum definition: connection with part implies connection with cover
part(E2,E1) :- space_cover_event(M,E1),
        has_event(M,E2).
```

Program 6.2: Construction of spatial cover event

Along with axiom 2 (all non-element regions are sums over path-connected sets) this then implies that the atomic regions in a single manifold must be path-connected. The associated program is not given here. As was the case for time, several conditions are needed to ensure spatial coherence of contents. Importantly, two different objects cannot occupy the same space at the same time:

Definition 37. *Spatial structure $\mathcal{S}$ is **coherent** when:*

1. *If the contents of events $e, f \in S$ represent distinct objects, then $\neg(SO(e,f) \wedge O(e,f))$. (where O represents temporal overlap).*
2. *If $e, f \in S$ have incompatible contents then $\neg(SO(e,f) \wedge O(e,f))$.*

Note that the spatial overlap is of relevance for the second condition: an object might be both light and dark, or hot and cold at the same time at *different regions* (for instance if a table is half in the shade). As a simple example I provide the FAE with a manifold of 25 input 'intuitions' with content $s(intensity, 1)$ without any distinction of moments. The FAE is constrained so that distinct sensations of the same property (and without any object) that are sensed at the same time must exist in different spaces. Thus, the result is a spatial structure of 25 atomic spaces. Figure 6.1 shows 3 examples of the found spatial structures $\mathcal{S}_U$. Note that 25 regions can be placed on a grid without inside corners only if the grid is 5×5 or 1×25. All atomic regions contain a sub-atomic region as proper part, and are covered by a region representing space as a whole.

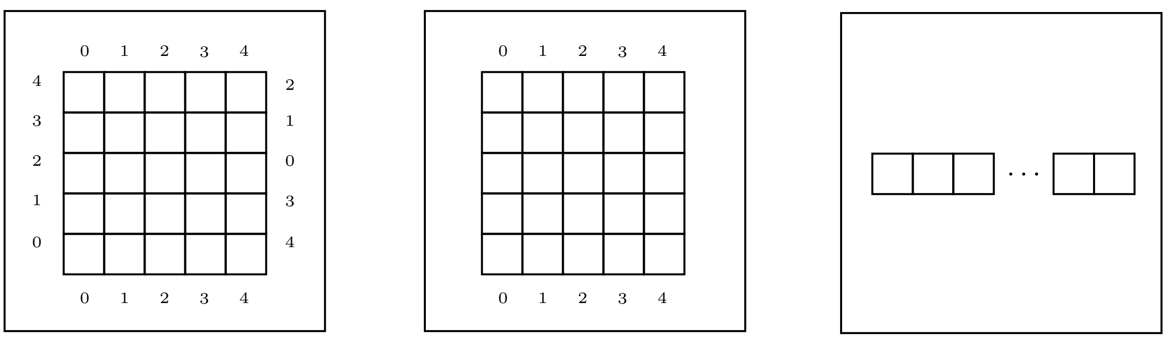


Figure 6.1: 3 spatial structures of 25 atomic regions. Numbers represent horizontal and vertical connections between the edges. The associated *tolerance spaces* are $ToK^2_{5,5}$, $TuK^0_{5,5}$ and $K_{25,1}$

.

6.3 Movement

Now to build a spatial structure encompassing *different manifolds*, the FAE is given 'motions' or 'movements' between positions "as action of the subject", associated with partial functions f, g, up etc.. Movements can be interpreted as *continuous functions* on the topological space of spaces defined in section 5.3: the inverse movement function applied to a space (open set) returns a space. Computationally, I define movements from spatial covers to spatial covers (i.e., sums over atomic regions). If there is no motion between two successive input positions, we let the spatial region of the two associated manifolds be equal. If there is a movement between two successive input indices, the two associated spatial regions are connected. Note that this ensures by the sum condition that at least two of the atomic regions contained in the manifolds are connected. Of course, there might even be an overlap, if the two larger spaces share atomic regions. I formulate coherence conditions for movements. The first four conditions provide general coherence. The latter three implement a specific interpretation of movement as traversal on a 2D plane:

1. Movements are not trivial: $f(A,B) \to \neg EQ(A,B)$.
2. Movements are invariant under EQ: $f(A,B) \wedge EQ(A,C) \to f(C,B)$, and $f(A,B) \wedge EQ(B,C) \to f(A,C)$.
3. Movements are injective functions: $f(A,B) \wedge f(A,C) \to EQ(B,C)$ and $f(A,B) \wedge f(C,B) \to EQ(A,C)$.
4. If there exists a sequence of moves $f = f_1 \circ f_2 .. \circ f_n$ such that $f(A,B)$ and $EQ(A,B)$, then for all regions C, D: $f(C,D)$ implies $EQ(C,D)$.
5. Movements are isomorphisms with respect to the connection relation C: $f(A,B) \wedge f(C,D)$ implies ($C(A,C)$ iff $C(B,D)$)
6. If $f(A,B)$ and A, B share x atomic regions, then the same number of atomic regions is shared between any C, D such that $f(C,D)$.
7. If $f(A,B)$ and there exist atomic regions $a\, b$ such that $P(a,A), P(b,B)$ and $x_i^a = b$ (i.e., the connection from a to b has label i), then for all C, D such that $f(C,D)$, there exist c, d such that $P(c,C), P(d,D)$ and $x_i^c = d$

I do not claim that this list of conditions is in any way exclusive. A full restrictions of movements to traversals on a 2D plane would require much more computationally expensive conditions such as a transitive variant of condition 7. The FAE now constructs a single spatial structure $\mathcal{S}_U$, using the same system of merging and embedding that was applied in the previous chapter. But now additionally the spaces that cover manifolds are connected in a manner that satisfies the conditions for movement given above, until at last the region u is constructed that covers all regions in $\mathcal{S}_u$. Again, this is our representation of space as a whole, resulting from the *figurative synthesis* in its pure form, and grounding the same synthesis in its empirical application. Embeddings are again identity functions from substructures to structures, so that the unified result equals the direct limit of the system. Importantly, all *cover* events are permanent representations of space itself, in accordance with what was said in chapter 2. A representation of the system of spatial embeddings is given in figure 6.2. Note that spatial covers of individual manifolds might very well have 'inside corners' or even gaps, as long as $\mathcal{S}_U$ satisfies the conditions of section 5.6.

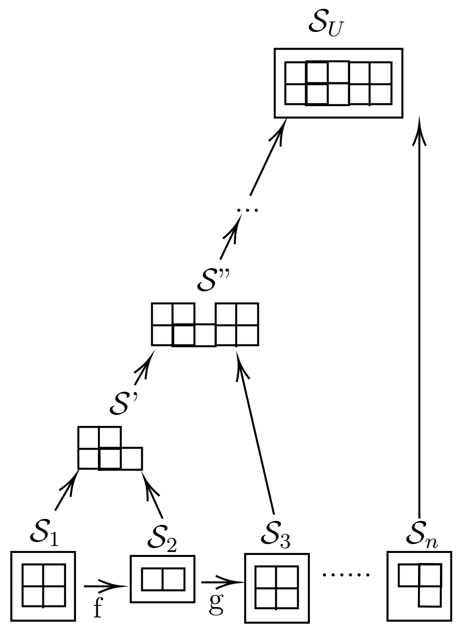


Figure 6.2: Structure of spatial embeddings in apprehension and reproduction. All covers are elements of $\mathcal{S}_U$, but this is not represented here to maintain simplicity in the diagram.

6.4 Making sense of spatial structures

Now again, I further complicate the definitions of *traces* and *making sense*:

Definition 38. *Given event structure $\mathcal{W}_T = (W_T, R_+, R_-, O, \preceq))$ and spatial structure $\mathcal{S}_T = (S_T, C)$, if $W_T = S_T$ there exists the following* ***spatio-temporal structure****: $\mathcal{WS}_T = (WS_T, R_+, R_-, O, \preceq, C)$, where $WS_T = W_T = S_T$.*

Definition 39. *Given theory $T = (\phi, I, R)$, the* ***spatio-temporal*** *trace of T is a spatio-temporal structure $\mathcal{WS}_T$ such that definitions 20 ($\mathcal{W}_T$ is a temporal trace) and 37 (spatial coherence) are satisfied and for $e, f \in WS_T$, if e, f represent the same object then $O(e, f) \rightarrow EQ(e, f)$ (objects are spatially coherent).*

Note that this final condition is not necessarily sensible for spatial structures in general and thus not part of the spatial coherence condition: intuitions at different spaces might represent the same object, as when a house is taken into view piece by piece. In the case of a trace however, we construct spatial events for *objects of experience*, so that these events make up the whole space of an object. A lot more may be done in specifying general conditions for the spatial trace of theories. For instance, one may insist that causes and effects must be connected (although this is not always sensible, e.g. magnetism). I however content myself with these minimal conditions for the moment, which suffice to explain a 'manifold of intuition' as a conceptual structure of objects in space.

Definition 40. *Theory T makes sense of spatiotemporal structure $\mathcal{WS}_U$, if it has a trace $\mathcal{WS}_T$ such that definition 21 (making temporal sense) is satisfied and:*

1. *For every event $e \in WS_U$ with objectless content $(p,(v_1,v_2))$, there exists event $f \in WS_T$ with content $(p,o,(v_3,v_4))$ so that $v_1 <= v_2$ implies $v_3 <= v_4$, $v_1 >= v_2$ implies $v_3 >= v_4$, the interval $[v_1,v_2]$ is enclosed in $[v_3,v_4]$ and $P(e,f)$.*

2. *For every event $e \in WS_U$ with content of the form $(o,(p,q))$, $(p,o,(v_1,v_2))$, of $(o1,o2,(p,q)$ there must exist event $f \in WS_T$ so that definition 21 is satisfied for e by f (i.e. f covers e in content as well as time), and $P(e,f)$.*

Note that, as in the previous chapter, $\mathcal{WS}_U$ and $\mathcal{WS}_T$ in fact constitute as single spatio-temporal structure. Hence, axiom 2 also applies: events in $\mathcal{WS}_T$ are sums over path-connected sets of atomic regions in $\mathcal{WS}_U$. This then ensures that the *objects of experience* introduced in $\mathcal{WS}_T$ exist in connected regions of space. Additional code is of course needed to implement these definitions. But this is a rather technical and perhaps unilluminating sequence of constraints that analyses contents and part relations. The main point of interest is that for atomic regions and events in $\mathcal{WS}_T$, the P relations are guessed explicitly instead of the connection relation C. This ensures that regions in $\mathcal{W}_T$ are sums over atomic regions, although additional constraints are implemented to implement path-connectedness.

```
nospace_overlap(A,E) :- atomic_space(A), concept_event(E),
    incompossible_events(A,E), time_overlap(A,E).

{ part(A,E) } :- concept_event(E), atomic_space(A),
    not nospace_overlap(A,E).
```

Program 6.3: Construction of part relation between atomic spaces and conceptual covers

6.5 Example behavior

We would like our system to adequately represent Kant's famous house example from the second analogy of experience. Kant explains that a house can often not be taken into view at once: "the apprehension of the manifold in the appearance of a house that stands before me is successive" (B236). However, the house itself is not perceived as a "happening". The key is that there is "no determinate order that made it necessary when I had to begin in the apprehension in order to combine the manifold empirically"(A193/B238). One might have started with the bottom of the house as well as its roof, since no *rule* determines an *objective sequence*. The example is taken up by Evans, who represents the house as a grid of cells with value 1 or 0. A computational agent can only view 2×2 blocks of this grid, and the AE is given a sequence of such blocks along with movements. The constructed theory then finds the correct assignment of 0 and 1 to the cells in the grid, along with the traversal associated with the movements. However, *space itself* is there to a large extent given. The cells, as well as there orientation in binary relations 'right' and 'below' are part of the input. Furthermore, the input specifies that all moves are distinct, and identifies the 2×2 input window as a set of 4 objects. A more *Kantian* approach is to start from intuition as manifold, which is apprehended in space. Only when synthesis produces unity in space and time can the *object of experience* be produced as conceptual counterpart to the manifold in intuition.

I apply the FAE to a simplified version of Evan's house (Evans, 2020). The FAE is given a sequence of inputs with value 0 or 1 as well as moves. I provide a simple template, without conceptual rules. The task for the FAE is to build a spatio-temporal structure and find a suitable interpretation for the moves. Note that inputs with the same content can be given more than once in the same position of the input ordering. The FAE then distinguishes the sensations in space, as when an agent receives light with the same intensity on different visual sensors.

Example 6.

$$\begin{aligned} S_1 &= \{intensity(0), intensity(0)\} \\ &move(right, (1,2)) \\ S_2 &= \{intensity(0), intensity(1)\} \\ &move(right, (2,3)) \\ S_3 &= \{intensity(0), intensity(1)\} \\ &move(left, (3,4)) \\ S_4 &= \{intensity(0), intensity(1)\} \\ &move(left, (4,5)) \\ S_5 &= \{intensity(0), intensity(0)\} \end{aligned}$$

$$\Phi = \begin{Bmatrix} T = \{cell, \}, \\ O = \{cell1, cell2, cell3 : cell4\} \\ P = \{intensity(x,0), intensity(x,1)\} \\ V = \{x : cell\} \end{Bmatrix} \begin{matrix} N_i = 4 \\ N_e = 0 \\ N_{RS} = 0 \\ N_{CR} = 0 \end{matrix}$$

Note that the atomic formulas in the input do not refer to any object. The unification of the manifold of intuition into objects is the result of constructing the trace $\mathcal{S}_T$. I provide a single constraint:

$$C = \{ \ \forall x : \ cell \ intensity(x,0) \ XOR \ intensity(x,1) \ \ \}$$

The FAE finds the following theory and spatial structures $\mathcal{S}_u$, $\mathcal{S}_T$, I represent the process of spatial synthesis through embeddings to show the underlying mechanism:

$$I = \{intensity(cell1,0), intensity(cell2,1) \ \ \}$$

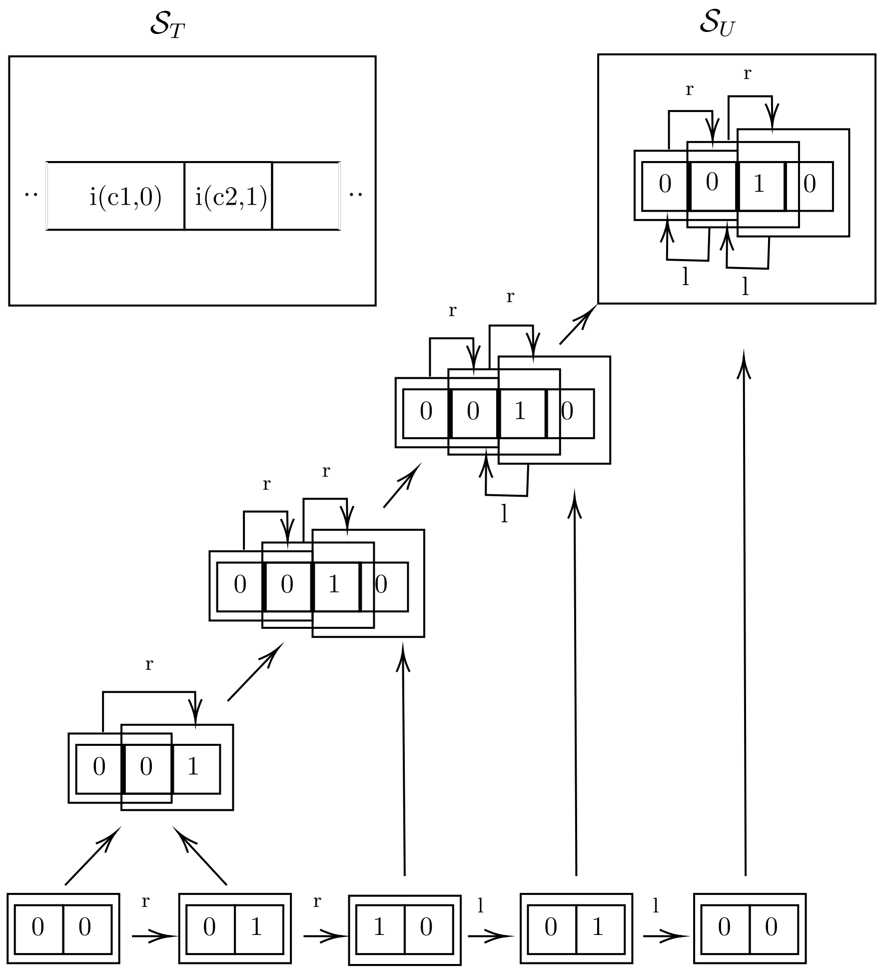


Figure 6.3: Spatial structures for example 6.

The constructed temporal structure is then rather uninteresting. Any structure that satisfies the overlap of events at the same input index and covering of the trace suffices" △

The FAE finds that the movements 'left' and 'right' are each others inverses. It also hypothesizes that both movements are associated with a spatial overlap: after moving to the right, a single atomic region that was perceived before is still in view. The spatial model constructed is that of a Turband, which allows the FAE to explain the sensory sequence with 2 instead of 3 objects. I now turn to an example where the spatial and temporal components of figurative synthesis are combined in a non-trivial manner. Kant discusses the example of *freezing water* as counterpart to the perception of a house:

"If (in another example) I perceive the freezing of water, I apprehend two states (of fluidity and solidity) as ones standing in a relation of time to each other. But in time, on which I ground the appearance as *inner intuition*, I represent necessary synthetic unity of the manifold, without which that relation could not be determinately given ... But now this synthetic unity, as the *a priori* condition under which I combine the manifold of an intuition in general, if I abstract from the constant form of my inner intuition, time, is the category of cause"

In this example the succession of perception thus corresponds to an *objective sequence* through causal unity. In the terminology of the second analogy: the order between the perception of water and that of ice is necessarily determined. One could not have perceived ice before water, in the same way that one could have perceived the roof of a house before its base by simply changing the direction of view. The change from water to ice happens in accordance with a causal rule.

Example 7. I again provide the FAE with perceived 'intensities' between which movements are made, but now the intensities change throughout the sensory sequence. The intended interpreatation is that the freezing of water results in higher light intensities. The task for the FAE is to decide whether it is perceiving multiple simultaneous objects or a single object changing through time in accordance with a causal rule.

$$\begin{aligned}
&S_1 = \{intensity(0)\}\\
&\quad move(right, (1, 2))\\
&S_2 = \{intensity(0), temp(sensor, 0)\}\\
&\quad move(left, (2, 3))\\
&S_3 = \{intensity(1)\}\\
&\quad move(right, (3, 4))\\
&S_4 = \{intensity(2)\}\\
&S_5 = \{intensity(3)\}
\end{aligned}$$

$$\Phi = \left\{ \begin{array}{l} T = \{cell, \}, \\ O = \{cell1, cell2, cell3 : cell\} \\ P = \{intensity(x, 0), intensity(x, 1), intensity(x, 2)\} \\ \qquad \cup \{temp(y, i) : 0 \leq i \leq 10\} \\ V = \{x : cell, y : sensor\} \end{array} \right\} \begin{array}{r} N_i = 3 \\ N_e = 6 \\ N_{RS} = 0 \\ N_{CR} = 1 \end{array}$$

I apply the same input constraint:

$$C = \Big\{\forall x : \ cell \ intensity(x, 1) \ XOR \ intensity(x, 1)[1ex] \quad \Big\}$$

The FAE finds a theory with a single cell, of which the light intensity rises as effect of the temperature being 0:

$$I = \Big\{intensity(cell1, 0), temp(sensor, 0)\Big\}$$

$$\text{R} = \Big\{(temp, y, (on, on) >>_{CR} (intensity, x, up)\Big\}$$

The found spatiotemporal structures are the following:

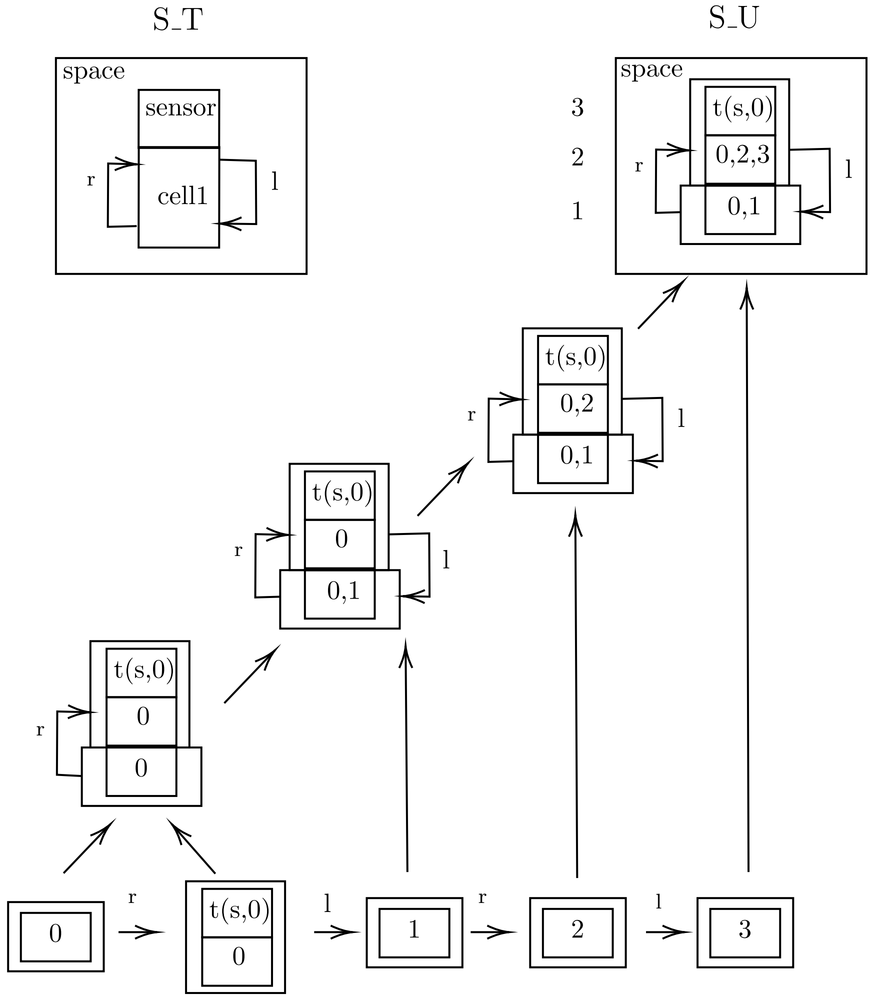


Figure 6.4: Spatial structure for example **x**. Comma's have been used to denote that several contents are associated with a single space throughout time. Numbering has been added to the atomic regions in $\mathcal{S}_U$ to signify the spatial component of events in the event structure below

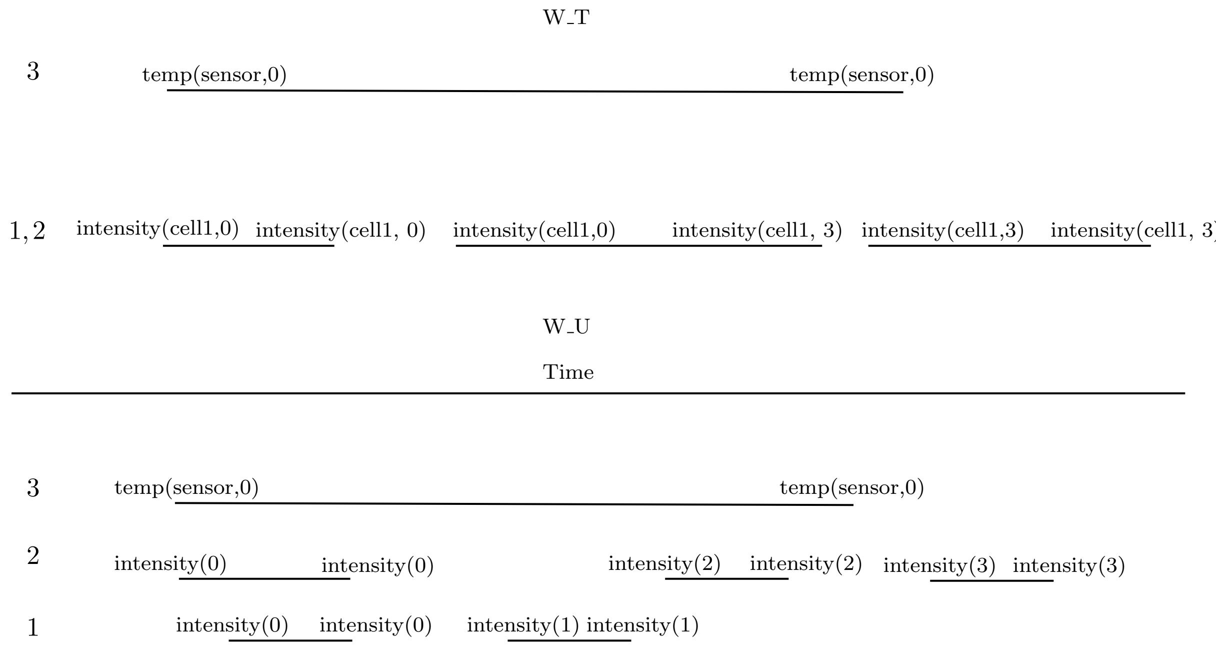


Figure 6.5: Temporal structure for example **x**. Indices on the left denote the atomic spaces in which events are placed

△

Here, the FAE learns that the movements 'right' and 'left' are each others inverses. Contrary to the previous examples, the movements are now not associated with a spatial overlap. The FAE hypothesizes that the two spaces between which the movements are made together contain a single object changing through time following a causal rule. As in Kant's example, the rule of change provides objective truth to the subjective sequence of 'intensities'. Contrary to the previous example, where the FAE finds multiple non-changing objects existing simultaneously, the sequence of appearances could not have been reversed by moving in the other direction.

Discussion and Future Research

In this thesis I have introduced several extensions of the AE. The implementation of geometric rules has been finished to a degree where application to non-trivial problems is possible. The implementation of *figurative synthesis* in space on the other hand is more a proof of concept: the conceptual and computational groundwork is laid for unification through *synthesis of apprehension*. In all cases more work can be done to generalize the developed frameworks and show their full potential. I have attempted to bridge the gap between the *Critique of Pure Reason* and the *Apperception Engine* simulatenously from a philosophical, mathematical and computational perspective. In this section I briefly go over the three previous chapters, pointing towards limitations of the implemented systems and directions for future research.

Geometric logic

In the third chapter I have introduced *geometric logic* as the logic of judgement and simultaneously the logic of *objective validity* on inverse systems. While both properties indicate the potential of this formalism, neither of them has been fully represented in the current implementation. On the one hand, I have only represented *hypothetical judgement*; the counterpart of causal rules in the AE. More functions from Kant's *table of judgements* must be cast in GEOLOG and provided as learnable structures to the AE before the subtle interaction of judgements that characterizes Kantian synthesis can take shape. On the other hand, the learned *geometric rules* are not interpreted on an inverse system, so that their *objective validity* remains implicit. Another limitation of the current approach is the restriction of *geometric implications* to *geometric rules* that do not allow for disjunctions in the consequents. Whether or not the full power of *geometric implications* is needed to adequately represent Kantian judgement is perhaps open to discussion. Although it is clear that Kant's disjunctive judgement and logical disjunctions are distinct operations, disjunctions might be useful in the subsumption of intuition under *sets of intensions*, see e.g. (Achourioti et al., 2017).

From a more practical point of view, the choice minimization process that favors precise theories over general theories must still be developed further. I have prevented 'fake' precise explanations of the form $number(y) \wedge equal(x, y) \rightarrow number(x)$ by requiring that the ground head of an explanatory rule may not be equal to one of its ground bodies, but this mechanism may falter when sufficiently large templates are provided. For example, in a set of rules $\{\exists x p(x), p(x) \rightarrow q(x), q(x) \rightarrow p(x)\}$, the third rule can still constitute a 'fake' precise explanation of $p(a)$. To resolve this, one might analyse the transitive closure of ground derivations within a single time, further constraining what counts as explaining a ground atom.

Making sense in time

I have shown how a sensory sequence can be represented as event structure through a process of embedding and merging. Time is then both *a priori* structure, and the result of a unifying process. Event structures represent the fundamental connectedness and potential infinity of Kantian time, and temporal synthesis underlies the synthesis of representation under concepts. While this extension then paves the way for a learning system that has fundamental similarities with Kant's architecture, the expressiveness of CONTENTLOG is still quite limited. I have not yet implemented the capacity to develop xor constraints. Furthermore, rules in CONTENTLOG have a single atom in their body since they represent relations between pairs of events and events have a single content. In order to generalize CONTENTLOG to a language that wholly includes DATALOG, we must either allow events to have multiple contents or define relations between more than two events. For instance, one might allow c_1 and c_2 to cause e *together* if and only if $c_1 \succeq e \wedge c_2 \succeq e$.

In relation to the *objective validity* of geometric logic, it seems that a more thorough model of *figurative synthesis* would apply retractions between event structures, instead of embeddings, and take $\mathcal{W}_U$ to be the inverse limit instead of the *direct limit*. In such a framework, *geometric sentences* could be constructed based on their truth over all event structures in the system of retractions. Their *transcendental truth* on the inverse limit would then be guaranteed by *objective validity*. Whether or not such an approach would come with computational as well as conceptual benefits will perhaps be made clear in the future. As discussed, constructing an inverse system of retractions in the current framework requires explicitly representing all causal pasts and futures, thereby cluttering the system. In the current approach we do have preservation of *geometric formulas* over embeddings. Interpreting implications as dynamical proofs (as discussed in chapter 3), we might then view *geometric formulas* as representations of 'what is' (objectively) that are built up throughout the embedding of event structures and only explained as a single system by implications on the conceptual level $\mathcal{W}_T$. In any case, more may be done to incorporate geometric logic and temporal construction in a single system and show how both extensions can complement one another.

Making sense in space

The implementation of spatial synthesis explores new grounds, and the identification of domain general inductive biasses that sufficiently constrain spatial construction has proven to be a hard nut to crack. More than the third and fourth chapter, the fifth and sixth chapters display a proof of concept, laying the mathematical and computational foundations for practical applications of spatio-temporal construction to be developed in future work. I have shown how *tolerance spaces* can be applied as a bridge between topological interpretations of the regional connection calculus and graph-theoretical dimensionality. Local conditions have proven sufficient to ensure a degree of global consistency for the spatial structure. As discussed, alternative topological structures might be applied in future research, potentially even structures that are *simply connected*. The current structure might also be extended to three-dimensional space, although the repercussions in terms of efficiency might be severe. From a more philosophical point of view, it would be interesting to analyse to what extent

RCC can be considered a *determination of sensibility by the understanding* as is the case for the temporal axiom system developped by Pinosio.

In the implementation of unification and *making sense* in space, much more may still be done. The same comments on geometric logic and embeddings that were given above apply to spatial construction. On the topic of spatial construction itself, an interesting avenue to pursue is to explore the relevance of motion to a higher degree. Pinpointing the topological properties that most adequately place subjective motion in the center of spatial synthesis might prove core in the development of a generalizable spatial synthesis system. Finally, the critical reader might have noted that the example of spatial synthesis given in this thesis display conceptual *explanation*, but no spatial *prediction*. Indeed, the interaction between concept formation, space and prediction is a topic that requires more intensive thought. Kant understood the *empirical concept* geometrically, stating that "the empirical concept of a plate has homogeneity with the pure geometrical concept of a circle" (B176). Future research might thus take up the spatial system developed in this work, and provide predicates with a geometric content in terms of the directed relations. Then, the unification of input under the concept 'house' by the FAE might result in predictions, as when one sees the bottom of a house and predicts that there must be a top. Note that such empirical concepts must be learned through experience. A proper implementation of *making geometric sense* might thus require a sequential implementation of the Apperception Engine in which achieved concepts can be stored and applied to later tasks.

Computational limitations

Throughout this project, a central theme and limitation has always been the limitations on running time and memory usage. Noting that the original AE already takes more than two days to perform some tasks, it is clear that computational bounds are of severe significance for any attempt to make the system more expressive. Especially the spatial construction introduced in chapters 5 and 6 provides a degree of freedom that renders application to larger examples such as Evans' house very difficult. Future approaches may abandon Clingo and move towards potentially more powerful program synthesis frameworks. One promising candidate is the 'Dream Coder' (Ellis et al., 2020), a program synthesis system that discovers explanatory symbolic structures in a machine learning framework. It would be very stimulating to discover the potential of a system that combines statistical learning with Kantian architecture.

Conclusion

As counterpart to the sizeable body of text that this thesis applies to convey its message, its concluding remarks contain no more words than are strictly necessary to repeat what has been said.

The Apperception Engine is a pioneering AI system in terms of explainability and generalizability. Still, Kant's architecture provides concepts and inspiration for many extensions of this system yet to come.

Non-deterministic choice over predefined domains can represent *geometric logic* and open the door to Kantian judgements as unifying functions and *rules*.

Taking *figurative synthesis* seriously means recognizing both *form of intuition* and *formal intuition* as determination of *sensibility* by the *understanding*.

By embedding qualitative structures, we may simultaneously unify intuition in time and space, represent time and space as boundless, unique and potentially infinite structures, provide conceptual reasoning with a spatio-temporal interpretation, and subsume intuition under *objects of experience*.

This project has further explored the potential for general application of artificial intelligence. The artificial agent that constructs his world as unity in space and time might one day interact with it as if it *experiences*, as if it *understands* its environment.

I thank the reader for her or his time and attention.

Appendix

This appendix contains a few interesting but relatively large programs. No explanation is provided, but the related section in the thesis is given.

4.5.3: Constraints on the temporal consistency between events

```
:- incompossible_events(E1, E2),
    time_overlap(E1, E2).

C1 = C3 :-
    time_overlap(E1, E2), has_content(E1, s(_,(C1,_))),
    has_content(E2, s(_,(C3,_))),
    same_object_and_property(E1,E1), r_after(E1, E2),
    r_after(E2, E1).

C2 = C4 :-
    time_overlap(E1, E2), has_content(E1, s(_,(_,C2))),
    has_content(E2, s(_,(_,C4))),
    same_object_and_property(E1,E1), r_before(E1, E2), r_before(E2, E1).

C2 = C4 :-
    time_overlap(E1, E2), has_content(E1, s2(_,_,(C1,C2))),
    has_content(E2, s2(_,_,(C3,C4))), events_same_obj_and_prop_s2(E1, E2),
    r_before(E1, E2), r_before(E2, E1).

C1 = C3 :-
    time_overlap(E1, E2), has_content(E1, s2(_,_,(C1,C2))),
    has_content(E2, s2(_,_,(C3,C4))), events_same_obj_and_prop_s2(E1, E2),
    r_after(E1, E2), r_after(E2, E1).


V3 >= V1 :-
    has_content(E1, s2v(C, O, (V1,V2))), V1 <= V2, has_content(E2, s2v(C, O, (
   V3,V4))), time_overlap(E1, E2), r_after(E2, E1).

V4 <= V2 :-
    has_content(E1, s2v(C, O, (V1,V2))), V1 <= V2,
    has_content(E2, s2v(C, O, (V3,V4))), time_overlap(E1, E2),
    r_before(E2, E1).

V3 <= V1 :-
    has_content(E1, s2v(C, O, (V1,V2))), V1 >= V2,
    has_content(E2, s2v(C, O, (V3,V4))), time_overlap(E1, E2),
    r_after(E2, E1).

V4 >= V2 :-
    has_content(E1, s2v(C,O,(V1,V2))), V1 >= V2,
```

```
    has_content(E2, s2v(C,O,(V3,V4))), time_overlap(E1, E2),
    r_before(E2, E1).


:- time_overlap(E1, E2), has_content(E1, s2v(C, O, (V1,V2))),
   has_content(E2, s2v(C, O, (V3,V4))), V1<V3, V1<V4,
   V2<V3, V2<V4.

:- time_overlap(E1, E2), has_content(E1, s2v(C, O, (V1,V2))),
   has_content(E2, s2v(C, O, (V3,V4))), V1<V2, V3>V4.
```

4.5.4: Merging manifolds and events

```
merged_manifold((1,D), (1, D-1), (D+1, 0)) :-
    manifold((1, D-1)), manifold((D+1, 0)).

has_event(A, Event) :-
    merged_manifold(A, B, _), has_event(B, Event).

has_event(A, Event) :-
    merged_manifold(A, _, C), has_event(C, Event).

has_event(A, E3) :-
    merged_manifold(A, B, C), has_event(B, E1),
    has_event(C, E2),
    merged_event(E3, E1, E2).

1 { merged_event(E1+X, E1, E2) : pos_intuit_event(E1+X) } 1 :-
    merged_manifold(A,B,C),
    events_same_object_and_property(E1, E2),
    input_size(X),
    last_of_its_kind(B, E1),
    first_of_its_kind(C, E2),
    not chain_same_value(E1,E2).

last_of_its_kind(M, E1) :-
    has_event(M, E1), not has_successor(M, E1),
    not has_more_precise_sim_ending_event(M, E1).

has_successor(M, E1) :-
    has_event(M, E1), has_event(M, E2),
    events_same_object_and_property(E1, E2),
    r_strictly_before(E1, E2).

has_more_precise_sim_ending_event(M, E1) :-
    has_event(M, E1), has_event(M, E2),
    events_same_object_and_property(E1, E2),
    r_before(E1, E2), time_covers(E1, E2),
    not time_covers(E2, E1).

first_of_its_kind(M, E1) :-
    has_event(M, E1), not has_predecessor(M, E1),
    not has_more_precise_sim_beginning_event(M, E1).
```

```
has_predecessor(M, E1) :-
    has_event(M, E1), has_event(M, E2),
    events_same_object_and_property(E1, E2),
    r_strictly_after(E1, E2).

has_more_precise_sim_beginning_event(M, E1) :-
    has_event(M, E1), has_event(M, E2),
    events_same_obj_and_prop_s(E1, E2),
    r_after(E1, E2), time_covers(E1, E2),
    not time_covers(E2, E1).

chain_same_value(E1, E2) :-
    has_content(E1, s(O,(_,C2))),
    has_content(E2, s(O,(C2,_))).

chain_same_value(E1, E2) :-
    has_content(E1, s2v(C, O, (_,V2))),
    has_content(E2, s2v(C, O, (V2,_))).
```

4.5.6: implementation of trace definition:

```
concept_event(X) :- pos_concept_event(X), has_content(X,_).

 %1: Init events from I

1{has_content(E,s2v(Concept, Object, (V,V))) : pos_init_event(E)}1 :-
                init(s2v(Concept, Object, V)).

1{has_content(E,s(Object, (Concept,Concept))) : pos_init_event(E)}1 :-
                init(s(Concept, Object)).

:- init_event(X), concept_event(Y),
    not r_after(Y,X).

 %2: Regular successions

1 {regular_succession(R,Subs,E1,E2) : pos_concept_event(E2), E1<E2}1:-
        concept_event(E1),
        is_regsucc_rule(R),
        rule_subs(R,Subs),
        eval_body(R,Subs,E1),
        not ending_concept_event(E1).

r_strictly_after(E2,E1) :- regular_succession(_,_,E1,E2).

:- regular_succession(_,_,E1,E2), time_overlap(E1,E2).

has_content(E2,GH) :-
        regular_succession(R,Subs,_,E2),
        rule_head_co(R,VH),
        rule_subs(R,Subs),
```

```
        ground_content(VH, GH, Subs).

has_dir_content(E2,s2v(C,Obj,up)) :-
        regular_succession(R,Subs,E1,E2), rule_head_co(R,s2v(C,Var,up)),
        rule_subs(R,Subs),
        subs(Subs,Var,Obj).

has_dir_content(E2,s2v(C,Obj,down)) :-
        regular_succession(R,Subs,E1,E2), rule_head_co(R,s2v(C,Var,down)),
        rule_subs(R,Subs),
        subs(Subs,Var,Obj).

:- regular_succession(_,_,E1,E2), not boundary_between(E1,E2).


 %3: Causal rules
1 {causes(R,Subs,E1,E2) : pos_concept_event(E2), E1 < E2}1:-
        concept_event(E1),
        is_causal_rule(R),
        rule_subs(R,Subs),
        eval_body(R,Subs,E1),
        not ending_concept_event(E1).

time_covers(E1,E2) :- causes(_,_,E1,E2).

has_content(E2,GH) :-
        causes(R,Subs,_,E2),
        rule_head_co(R,VH),
        rule_subs(R,Subs),
        ground_content(VH, GH, Subs).

has_dir_content(E2,s2v(C,Obj,up)) :-
        causes(R,Subs,E1,E2), rule_head_co(R,s2v(C,Var,up)),
        rule_subs(R,Subs),
        subs(Subs,Var,Obj).


has_dir_content(E2,s2v(C,Obj,down)) :-
        causes(R,Subs,E1,E2), rule_head_co(R,s2v(C,Var,down)),
        rule_subs(R,Subs),
        subs(Subs,Var,Obj).


 %4: Stabilisation for alteration events.
1 {stabilises(s(O,(C1,C2)),E1,E2) : pos_concept_event(E2), E1 < E2}1 :-
        concept_event(E1),
        has_content(E1, s(O,(C1,C2))),
        C1 != C2.


1 {stabilises(s2v(C,O,(V1,V2)),E1,E2) : pos_concept_event(E2), E1<E2}1 :-
        concept_event(E1),
        has_content(E1, s2v(C,O,(V1,V2))),
        V1 != V2.
```

```
has_content(E2, s(O,(C2,C2))) :- stabilises(s(O,(C1,C2)), _,E2).
has_content(E2, s2v(C,O,(V2,V2))) :- stabilises(s2v(C,O,(V1,V2)), _,E2).

:- stabilises(_,_,E1,E2), not boundary_between(E1,E2).

:- stabilises(_,E1,E2), r_before(E2,E1).

:- stabilises(_,E1,E2), time_overlap(E1,E2).


 %5 Minimality follows from generation of events under 1-4

 %6: Every non-init event has a unique predecessor as starting point.
has_prev(E2) :-     boundary_between(E1,E2), chain_same_value(E1,E2).

:- concept_event(E2), not init_event(E2), not has_prev(E2).

:- has_prev(E1, E2), has_prev(E3, E2), E1 != E3.

:- has_prev(E1, E2), has_prev(E1, E3), E2 != E3.

 %7 No ambiguity

:- time_overlap(E1,E2), concept_event(E1), concept_event(E2),
has_content(E1, s(_,(C1,C2))), has_content(E2, s(_,(C3,C4))),
events_same_obj_and_prop_s(E1,E2),
r_strictly_before(E2,E1), C4 != C1, C4 != C2.

:- time_overlap(E1,E2), concept_event(E1), concept_event(E2),
has_content(E1, s(_,(C1,C2))), has_content(E2, s(_,(C3,C4))),
events_same_obj_and_prop_s(E1,E2),
r_strictly_after(E2,E1), C3 != C1, C3 != C2.

% 8 No covering of previous.

:- causes(_,_,E1,E2), has_dir_content(E2, s2v(C,O,D1)),
     r_before(E3,E2), has_dir_content(E3, s2v(C,O,D2)),
     D1 != D2, time_overlap(E1,E3).

:- causes(R,_,E1,E2), concept_event(E3),
     events_same_obj_and_prop_s(E2,E3),
     E2 != E3, r_before(E3,E2),
     time_covers(E1,E3).
```

6.2; Implementation of locally sub-King and symmetry consistency conditions, as well as prohibition of inside corners.

```
neighbour_number(0..7).

even(0;2;4;6).
odd(1;3;5;7).
```

```
nextto(I, I+1) :- neighbour_number(I), neighbour_number(I+1).
nextto(7,0).
nextto(I,I) :- neighbour_number(I).
nextto(I1,I2) :- nextto(I2,I1).
nextto(0,2).
nextto(2,4).
nextto(4,6).
nextto(6,0).

space_designator(E1) :- atomic_space(E1), E1 <= E2 : space_equal(E1,E2).

1{neighbour(E1,E2,I) : neighbour_number(I)}1 :-
        atomic_space(E1), atomic_space(E2), connected(E1, E2), not space_equal
    (E1,E2), space_designator(E1), space_designator(E2).

:- neighbour(N,E1,I), neighbour(N,E2,I), E1 != E2.

:- neighbour(N, E1, I1), neighbour(N, E2, I2), nextto(I1,I2), not connected(E1
    , E2).
:- neighbour(N, E1, I1), neighbour(N, E2, I2), not nextto(I1,I2), connected(E1
    , E2).


:- locally_grid, neighbour(A,N1,I1),
even(I1), neighbour(A,N2,I2), I2 = (I1+2)\8, neighbour(N1,Third,_),
neighbour(N2,Third,_),
        A != Third, not neighbour(A,Third,I1+1).

:- locally_grid, neighbour(A,N1,I1),
even(I1), neighbour(A,N2,I2), I2 = (I1+4)\8,
neighbour(N1,Third,_), neighbour(N2,Third,_),
        A != Third,  not neighbour(A, Third, I3), not neighbour(A, Third, I4),
     I3 = (I1+2)\8, I4 = (I1+6)\8.

saturated(A) :-
    space_designator(A),
        neighbour(A,_,I): neighbour_number(I).

locally_grid :-
    saturated(A) : space_designator(A).

% symmetry consistency
neighbour(E2,E1,I2) :- neighbour(E1,E2,I1), I2 = (I1+4)\8.

% No inside corners
:- neighbour(E,_,I), odd(I), J = (I+1)\8, not neighbour(E,_,J).
:- neighbour(E,_,I), odd(I), J = (I-1)\8, not neighbour(E,_,J).

:- neighbour(E,_,I), even(I), J = (I+2)\8, neighbour(E,_,J),
        L = (I+1)\8, not neighbour(E,_,L).
```

References

Abiteboul, S., Hull, R., & Vianu, V. (1995). *Foundations of Databases* (Vol. 8). Addison-Wesley Reading.

Achourioti, T., & van Lambalgen, M. (2011). A Formalization of Kant's Transcendental Logic. *The Review of Symbolic Logic*, *4*(2), 254–289.

Achourioti, T., van Lambalgen, M., et al. (2017). Kant's Logic Revisited. *IfCoLog Journal of Logics and their Applications*, *4*, 845–865.

Aiello, M., Pratt-Hartmann, I., Van Benthem, J., et al. (2007). *Handbook of Spatial Logics* (Vol. 4). Springer.

Alexandroff, P. (1937). Diskrete Raume. *Mat. Sbornik, Number 3*, *2(44)*(3), 501–519.

Blass, A. (1993). Topoi and Computation. In *Current Trends In Theoretical Computer Science: Essays and Tutorials* (pp. 310–317). World Scientific.

Chen, Z.-Z., Grigni, M., & Papadimitriou, C. H. (2002). Map Graphs. *Journal of the ACM*, *49*(2), 127–138.

Cohn, A. G., Bennett, B., Gooday, J., & Gotts, N. M. (1997). Qualitative Spatial Representation and Reasoning with the Region Connection Calculus. *GeoInformatica*, *1*(3), 275–316.

Coquand, T. (2010). A Completeness Proof for Geometric Logic. In *Logic, Methodology and Philosophy of Science. Proceedings of the Twelfth International Congress* (pp. 79–90).

Dantsin, E., Eiter, T., Gottlob, G., & Voronkov, A. (1997). Complexity and Expressive Power of Logic Programming. In *Proceedings of Computational Complexity. Twelfth Annual IEEE Conference* (pp. 82–101).

Dennett, D. C. (2018). From Bacteria to Bach and Back: The Evolution of Minds. *Evolutionary Studies in Imaginative Culture*, *2*(1), 113–116.

Dornheim, C. (1998). Undecidability of Plane Polygonal Mereotopology. *KR*, *98*, 342–353.

Dyckhoff, R., & Negri, S. (2015). Geometrisation of First-Order Logic. *Bulletin of Symbolic Logic*, *21*(2), 123–163.

Eiter, T., Faber, W., Fink, M., & Woltran, S. (2007). Complexity Results for Answer Set Programming with Bounded Predicate Arities and Implications. *Annals of Mathematics and Artificial Intelligence*, *51*(2), 123–165.

Eiter, T., Leone, N., & Sacca, D. (1998). Expressive Power and Complexity of Partial Models for Disjunctive Deductive Databases. *Theoretical Computer Science*, *206*(1-2), 181–218.

Ellis, K., Wong, C., Nye, M., Sable-Meyer, M., Cary, L., Morales, L., ... Tenenbaum, J. B. (2020). Dreamcoder: Growing Generalizable Interpretable Knowledge with Wake-Sleep Bayesian Program Learning. *arXiv preprint arXiv:2006.08381*.

Evans, R. (2020). *Kant's cognitive architecture* (Unpublished doctoral dissertation). Imperial College London.

Falkenstein, L. (2018). *Kant's Intuitionism.* University of Toronto Press.

Fisher, J., & Bezem, M. (2007). Skolem Machines and Geometric Logic. In *International Colloquium on Theoretical Aspects of Computing* (pp. 201–215).

Gao, P., Gu, A., & Zakhor, A. (2008). Optical Proximity Correction with Principal Component Regression. In *Optical Microlithography XXI* (Vol. 6924, p. 69243N).

Gebser, M., Kaminski, R., Kaufmann, B., & Schaub, T. (2012). *Answer Set Solving in Practice*. Morgan & Claypool Publishers.

Giannotti, F., Pedreschi, D., Sacca, D., & Zaniolo, C. (1991). Non-Determinism in Deductive Databases. In *International Conference on Deductive and Object-Oriented Databases* (pp. 129–146).

Greco, S. (1998). Non-Determinism and Weak Constraints in Datalog. *New Generation Computing*, *16*(4), 373–395.

Grigni, M., Papadias, D., & Papadimitriou, C. (1995). Topological Inference. In *IJCAI (1))* (pp. 901–907).

Henkin, L., Suppes, P., & Tarski, A. (1959). *The Axiomatic Method*. North-Holland Amsterdam.

Hodges, W. (1993). *Model Theory*. Cambridge University Press.

Kant, I. (1966). *Prolegomena to Any Future Metaphysics That Will Be Able to Come Forward as Science*. Manchester University Press.

Kant, I. (1998). *Critique of Pure Reason*. New York, NY: Cambridge University Press. (Translated by Paul Guyer and Allen W. Wood)

Kant, I. (2001). *Lectures on Metaphysics*. Cambridge University Press.

Legg, S., & Hutter, M. (2007). Universal Intelligence: A Definition of Machine Intelligence. *Minds and machines*, *17*(4), 391–444.

Li, S., & Ying, M. (2003). Region Connection Calculus: Its models and composition table. *Artificial Intelligence*, *145*(1-2), 121–146.

Ligozat, G. (2013). *Qualitative Spatial and Temporal Reasoning*. John Wiley & Sons.

Longuenesse, B. (2020). *Kant and the Capacity to Judge: Sensibility and discursivity in the transcendental analytic of the Critique of Pure Reason*. Princeton University Press.

Marcus, G. (2018). Deep Learning: A Critical Appraisal. *arXiv preprint arXiv:1801.00631*.

Márquez, A., de Mier, A., Noy, M., & Revuelta, M. (2003). Locally Grid Graphs: Classification and Tutte Uniqueness. *Discrete mathematics*, *266*(1-3), 327–352.

Palmgren, E. (2002). An Intuitionistic Axiomatisation of Real Closed Fields. *Mathematical Logic Quarterly: Mathematical Logic Quarterly*, *48*(2), 297–299.

Pinosio, R. (2017). *The Logic of Kant's Temporal Continuum* (Unpublished doctoral dissertation). University of Amsterdam.

Randell, D. A., Cui, Z., & Cohn, A. G. (1992). A Spatial Logic based on Regions and Connection. *KR*, *92*, 165–176.

Rossman, B. (2008). Homomorphism Preservation Theorems. *Journal of the ACM*, *55*(3), 1–53.

Sossinsky, A. B. (1986). Tolerance Space Theory and Some Applications. *Acta Applicandae Mathematica*, *5*(2), 137–167.

Vickers, S. (1993). Geometric Logic in Computer science. In *Theory and Formal Methods* (pp. 37–54). Springer.