Topos Theory as a Framework for Partial Truth

J. Butterfield

All Souls College
Oxford OX1 4AL
email: jb56@cus.cam.ac.uk

16 January 2000

Abstract

This paper develops some ideas from previous work (coauthored, mostly with C.J.Isham). In that work, the main proposal is to assign as the value of a physical quantity in quantum theory (or classical physics), not a real number, but a certain kind of set (a sieve) of quantities that are functions of the given quantity. The motivation was in part physical—such a valuation illuminates the Kochen-Specker theorem; in part mathematical—the valuations arise naturally in the theory of presheaves; and in part conceptual—the valuations arise from applying to propositions about the values of physical quantities some general axioms governing partial truth for any kind of proposition.

In this paper, I give another conceptual motivation for the proposal. I develop (in Sections 2 and 3) the notion of a topos (of which presheaves give just one kind of example); and explain how this notion gives a satisfactory general framework for making sense of the idea of partial truth. Then I review (in Section 4) how our proposal applies this framework to the case of physical theories.

To appear in *Proceedings of the 11th International Congress of Logic, Methodology and Philosophy of Science*, edited by Peter Gardenfors, Katarzyna Kijania-Placek and Jan Wolenski, Kluwer Academic.

1 Introduction

This paper develops some ideas from previous work (coauthored, mostly with C.J.Isham).[1] In that work, the main proposal is to assign as the value of a physical quantity in quantum theory (or classical physics), not a real number, but a certain kind of set (a sieve) of quantities that are functions of the given quantity.

In Isham and Butterfield (1998), we motivated this proposal in terms of (i) its relevance to foundations of quantum theory—such a valuation illuminates the Kochen-Specker theorem; and (ii) its mathematical naturalness—the valuations arise naturally in the theory of presheaves, which are a natural generalization of the notion of a set. Roughly speaking, a presheaf is a set whose membership varies from one context to another; and in that paper, a context was given by a quantum-theoretic quantity (a bounded self-adjoint operator). In Butterfield and Isham (1999), we developed some conceptual motivations for the proposal. We explained how the proposed valuations arise much more generally than just for quantum physics. In particular, they arise just as naturally in classical physics; and more generally, they arise from applying to propositions about the values of physical quantities some general axioms governing partial truth for any kind of proposition. In Hamilton, Isham and Butterfield (1999), we developed the proposal for a still more general notion of context: a context was taken, not as a physical quantity, but as a commutative von Neumann algebra. In all three papers, an analogue for sieve-valued valuations of the famous *FUNC* condition, familiar from discussions of the Kochen-Specker theorem, played a central role.

The aim of this paper is to give another conceptual motivation for the proposal, which again relates to the idea of partial truth. Our previous papers discussed partial truth entirely within the framework of presheaves (even in the most general Sections of Butterfield and Isham (1999)). But presheaves provide just one kind of example, among many, of the mathematical notion of a topos. A topos is a sort of category. Very roughly, it is a category that behaves much like the category of sets; indeed, this category, which I will call $\mathcal{S}$, is itself a topos. The exact definition of a topos is complicated; but in fact, I will not need it in this paper. For my purposes, the main point is that in general—not just for presheaves—the notion of a topos provides a good framework for making sense of the idea of partial truth. Thus while $\mathcal{S}$ has just the two usual truth-values, $\{0, 1\}$, a topos in general has a larger collection of truth-values. And just as $\{0, 1\}$ is itself a set, i.e. an object in the category $\mathcal{S}$, so also in any topos the collection of truth-values is an object in the topos: it is called 'the subobject classifier', and written Ω.

This point—that in any topos, one can make good sense of partial truth—is well-known to logicians working in category theory and topos theory. But it is largely unknown among philosophers and physicists, even those sympathetic to the idea of somehow 'going beyond' the usual classical two truth-values. So it seems worthwhile to expound it; and all the more worthwhile to do so using two simple examples of toposes (viz. the category of endomaps, and the category of graphs) that provide simple analogues of the previous papers' proposal. I undertake this project in Sections 2, 3 and 4; which discuss, respectively, categories in general, toposes in general, and our proposal—to define partial truth-values in quantum theory and other physical theories, using toposes of presheaves.

I should stress at the outset four features of this framework for partial truth. All but the second are features that apply in any topos.

1. First, one common reason for being suspicious of notions of partial truth, and of many-valued logics, is the idea that the proposed notions are arbitrary: that other definitions,

e.g. about how to define truth-functional connectives, would be just as well motivated by the logico-semantic phenomena or intuitions appealed to, as are the proposed definitions. But this reason does not apply here. For as we shall see, in any topos, the collection of generalized truth-values (the sub-object classifier Ω) is completely fixed by the structure of the topos; (and in general, there is no other topos with a similar but different Ω that one can argue to be as well motivated as the given one). In short, the many-valued logics that arise in toposes are *not* arbitrary.

2. In the toposes I shall be concerned with (viz. the topos of endomaps, the topos of graphs, and toposes of presheaves) the generalized truth-values (the elements of Ω) are natural in another sense: they are the answers to a natural 'multiple-choice question' about the objects in the topos. Besides, it will be a question whose answers exhibit a neat meshing between the various degrees or kinds of partial truth, and total truth; (where total truth can be roughly identified with classical truth, $1 \in \{0,1\}$, and with the designated truth-value of traditional many-valued logics). For example, for graphs, the question will be, roughly speaking, 'What part of the given arc (i.e. arrow in a graph) lies within the given subgraph?'; and the answer will be closer to total truth, the more of the given arc is in the subgraph. Furthermore, in toposes of presheaves (the topic of our main proposal), this meshing is even stronger: the partial truth-values are completely controlled by the notion of total truth (in a precise sense discussed in Butterfield and Isham (1999)).

   So much by way of addressing general suspicion of notions of partial truth. The next two features are more formal; (and for reasons given at the start of Section 2, I will not discuss them in detail in this paper).

3. Any topos has notions of sum and product, which generalize the notions familiar from arithmetic and logic; with the sum and product obeying the familiar distributive law. So in relation to quantum theory, our proposal's retention of the distributive law sets us apart from the dominant tradition of quantum logic. (I will not here try to defend this departure from tradition; but only note that it has been argued, against this tradition, that the distributive law is indispensible to genuinely logical reasoning.)

4. On the other hand, our proposals do involve non-Boolean structure. More precisely: Any topos allows the definition of logical operations of 'and', 'or' and 'not' on its set of truth-values, in such a way that they obey an intuitionistic logic. For some toposes, like the category of sets, they also obey the other principles of classical logic; but *not* for our proposal's toposes of presheaves.

2 Category Theory

In this Section and the next, I develop those aspects of the notion of a topos, that show how it provides a framework for discussing partial truth. More specifically, in Section 3 I will present the clause in the definition of a topos that requires the existence of the subobject classifier, Ω. To understand this clause, and how Ω functions as a collection of generalized truth-values, requires familiarity with some other notions of category theory: among them, the notions of monic, terminal object, point (also called 'global element') and generalized element. So in this Section, I first introduce the idea of a category and my examples of categories, and then present these notions (in the order listed).

Two warnings about the limitations of what follows. (1) A full understanding of 'how Ω functions' of course requires the full definition of a topos, which I cannot give here, for reasons of space. For the definition includes clauses requiring the existence for any pair of objects A, B in the category, of certain other objects, such as a product $A \times B$ and an exponential A^B; where these notions generalize the set-theoretic notions, and are characterized wholly in terms of the existence and uniqueness of certain arrows (morphisms) between the objects—but also take some space to state! But fortunately, we can see a good deal about 'how Ω functions', without these clauses. More specifically, they ensure that the generalized truth-values have an algebraic structure so as to obey an intuitionistic logic, including the distributive law; (as discussed in items 3. and 4. of Section 1). But even while neglecting this algebraic and logical structure, we will be able to see clearly how Ω encodes different degrees of partial truth.

(2) Space limitations also prevent my giving proofs, but they are all elementary and in the standard textbooks. I also cannot give many examples of categories (or diagrams). But I will take as examples, in addition to $\mathcal{S}$ and categories of presheaves, the category of endomaps and the category of graphs. These are good examples not just because they are toposes. As we shall see, their subobject classifiers (i) can be drawn vividly—for example, in the category of graphs, the object Ω is itself a graph; and (ii) provide simple analogues of the (perhaps off-puttingly complicated!) subobject classifier in a topos of presheaves.[2]

So I shall present in the following subsections: (1) the definition of a category, and my examples other than toposes of presheaves; (2) monics (presented as part of a discussion of inverses); (3) terminal objects, points and generalized elements.

2.1 Categories; Examples

A category consists of a collection of *objects* and a collection of *arrows* (or *morphisms*), with the following three properties. (1) Each arrow f is associated with a pair of objects, known as its 'domain' (dom f) and the 'codomain' (cod f), and is written in the form $f : B \to A$ where $B = \mathrm{dom} f$ and $A = \mathrm{cod} f$. (2) Given two arrows $f : B \to A$ and $g : C \to B$ (so that the codomain of g is equal to the domain of f), there is a composite arrow $f \circ g : C \to A$; and this composition of arrows obeys the associative law. (3) Each object A has an identity arrow, $id_A : A \to A$, with the properties that for all $f : B \to A$ and all $g : A \to C$, $id_A \circ f = f$ and $g \circ id_A = g$.

I have already mentioned the prototype category (indeed, topos) $\mathcal{S}$, in which the objects are sets and the arrows are ordinary functions (set-maps) between them. In many categories, the objects are sets equipped with some type of additional structure, and the arrows are functions that preserve this structure. An obvious algebraic example is the category of groups, where an object is a group, and an arrow $f : G_1 \to G_2$ is a group homomorphism from G_1 to G_2. More generally, one often defines one category in terms of another; and in such a case, there is often only one obvious way of defining composition and identity arrows for the new category.

In addition to the category $\mathcal{S}$ of sets, I take as examples (both of which are toposes):

1. The category $\mathcal{E}$ of endomaps of sets. That is: starting with the category $\mathcal{S}$ of sets, with functions between sets as arrows, one defines:
   (i) an object of $\mathcal{E}$ is to be a set X equipped with an endomap α (*i.e.* α is an ordinary function on X, $\alpha : X \to X$); I write such an object as $(X; \alpha)$;
   (ii) an arrow of $\mathcal{E}$ is to be a function that preserves the endomap structure, *i.e.* an arrow from $(X; \alpha)$ to $(Y; \beta)$ is a function $f : X \to Y$ such that $f \circ \alpha = \beta \circ f$;

and then one shows that with the obvious definitions of composition and of identity functions, the associative and identity laws hold.

2. The category $\mathcal{G}$ of (irreflexive, directed) graphs.
(i) An object is a pair of sets, R and D, equipped with a pair of functions, $s : R \to D$ and $t : R \to D$, where R is called the set of *arcs* (or *arrows*) of the graph, D the set of *dots* of the graph, and s and t are called the *source* and *target* functions. So for an arc $r \in R$, $s(r) \in D$ is the dot that is the source of r, and $t(r)$ is its target. I shall write an object as $(s, t : R \to D)$.
(ii) Intuitively, an arrow of $\mathcal{G}$ (not to be confused with an arc in a graph!) is to be a function that 'preserves the graph structure', mapping arcs to arcs and dots to dots in such a way as to preserve source and target relations between arcs and dots. That is, an arrow from $(s, t : R \to D)$ to $(s', t' : R' \to D')$ is a pair of functions $f_{Ar} : R \to R'$ and $f_{Do} : D \to D'$, for which both:
$$f_{Do} \circ s = s' \circ f_{Ar}; \text{ and } f_{Do} \circ t = t' \circ f_{Ar}. \tag{2.1}$$
And again one shows that with the obvious definitions of composition and of identity functions, the associative and identity laws hold.

2.2 Inverses, monics and epics

In any category, an arrow $f : A \to B$ is called *invertible* (or *an isomorphism*, or *iso*) iff there is an arrow $g : B \to A$ such that $g \circ f = id_A$ and $f \circ g = id_B$. Such a g is unique, and is called the *inverse* of f, written f^{-1}; and then A and B are called *isomorphic*. In $\mathcal{S}$, objects i.e. sets are isomorphic iff they have the same number of elements; in $\mathcal{E}$ and $\mathcal{G}$, objects are isomorphic iff they 'have the same diagram' (in a sense that can straightforwardly be made precise in terms of functions). If f has an inverse, then f satisfies:

$$\text{If } f \circ h = f \circ k, \text{then } h = k; \text{'left cancellation'}; \tag{2.2}$$

$$\text{If } h \circ f = k \circ f, \text{then } h = k; \text{'right cancellation'}. \tag{2.3}$$

I turn to the notion of a monic, and the corresponding ('dual') notion of an epic; these are defined in terms of the cancellation laws. By way of motivation, I will also note the connection to 'one-sided inverses', i.e. retractions and sections.[3]

An arrow $f : A \to B$ is called *injective for arrows from* T iff for any pair of arrows $x_1 : T \to A$ and $x_2 : T \to A$, if $f \circ x_1 = f \circ x_2$ then $x_1 = x_2$. If f is injective for arrows from T to A, for every T, we say f is *injective*, or a *monomorphism*, or a *monic*. In other words, f is called monic iff f obeys left-cancellation, Eq. (2.2).

In $\mathcal{S}$, this definition is equivalent to the usual meaning of 'injective' as the function being one-to-one. In $\mathcal{E}$, with objects $(X; \alpha)$, the monics from $(X; \alpha)$ to $(Y; \beta)$ are the arrows, i.e. functions from X to Y, that (i) are one-to-one as functions and (ii) have a range closed under β. (Requirement (ii) follows just from the definition of an arrow in $\mathcal{E}$.) Similarly in $\mathcal{G}$, with objects $(s, t : R \to D)$: a monic is a pair of functions, f_{Ar} on the arrows and f_{Do} on the dots, that are each one-to-one as functions; where again, the requirement that the pair preserve the structure of the graph that is the domain of the arrow follows just from the pair's being an arrow in $\mathcal{G}$. Thus the 'copy' of the domain-graph $A = (s, t : R \to D)$ that lies within the codomain-graph

$B = (s', t' : R' \to D')$ can have more structure than A. It can have more arcs between 'a' pair of dots than A does, and it need not be a disconnected component of B.

A simple sufficient condition for being monic is given by the notion of a retraction. Given $f : A \to B$, a *retraction* for f is a 'left-inverse' of f, i.e. an arrow $r : B \to A$ such that $r \circ f = id_A$. It follows that if $f : A \to B$ has a retraction, then f is a monic.

In $\mathcal{S}$, the converse of this result is 'almost true' in that the only exception involves the empty set as domain. That is to say, if $f : A \to B$ is an injective function, and $A \neq \emptyset$, then f has a retraction, i.e. there are maps $r : B \to A$ such that $r \circ f = id_A$. (So the restriction to A non-empty is due to the fact that while there is a function, called the empty function, with domain $\emptyset$ and any codomain, there is no function with non-empty domain and $\emptyset$ as its codomain.)

But in a general category, the converse is not true: having a retraction is stronger than being monic. The simplest counterexample is a category with just two objects, say A and B, and one arrow between them, say from A to B (apart from the identity arrows on A and on B). This arrow is monic (for only id_A can be 'done before' it), but has no retraction. (Exercise! Think of a simple counterexample in $\mathcal{E}$ or in $\mathcal{G}$.)

The corresponding ('dual') definitions and result for surjectivity, rather than injectivity, are as follows. An arrow $f : A \to B$ is called *surjective for arrows to* T iff for any pair of arrows $x_1 : B \to T$ and $x_2 : B \to T$, if $x_1 \circ f = x_2 \circ f$ then $x_1 = x_2$. If f is surjective for arrows from B to T, for every T, we say f is *surjective*, or an *epimorphism* or *an epic*. In other words, f is called epic if f obeys right-cancellation, Eq (2.3). In $\mathcal{S}$, this definition is equivalent to the usual meaning of 'surjective'. (Exercise: Characterize the epics in $\mathcal{E}$ or in $\mathcal{G}$.)

A simple sufficient condition for being epic is given by the notion of a section. Given $f : A \to B$, a *section* for f is a 'right-inverse', i.e. a map $s : B \to A$ such that $f \circ s = id_B$. It follows that if $f : A \to B$ has a section, then f is an epic.

As before, the converse of this result is 'almost true' in $\mathcal{S}$, in that the only exception involves the empty set. But in a general category, the converse is not true: having a section is stronger than being epic. The simple example above will do: the category with just two objects, say A and B, and one arrow between them, say from A to B (apart from the identity arrows). This arrow is epic (for only id_B can be 'done after' it), but has no section.

Finally, on the relation between retractions and sections: If f has both a retraction r and a section s, then $r = s = f^{-1}$. But f can be both monic and epic, without being iso. Again, the simplest counterexample is the category with just two objects, say A and B, and one arrow between them, say from A to B.

2.3 Terminal Objects, Points and Generalized Elements

In this Subsection, I will first discuss a type of object and a type of arrow that may exist in a category: viz., terminal objects and points (also called global elements), illustrating them in $\mathcal{S}$, $\mathcal{E}$ and $\mathcal{G}$. This will lead to my last piece of stage-setting for Section 3's presentation of toposes, the notion of a generalized element; though as I mentioned, a full presentation of toposes would also require the notions of product and exponential (also called map object).

An object T is called a *terminal object* if for every object A there is exactly one arrow $f : A \to T$. We sometimes stress the uniqueness of the arrow using !, by writing $f : A \xrightarrow{!} T$. Any two terminal objects are isomorphic. We fix on one such object and write it as $\mathbf{1}$. (Dually, an object S is called an *initial object* if for every object A there is exactly one arrow $f : S \to A$;

initial objects are all isomorphic, and we fix on one such and write it as $\mathbf{0}$.) An arrow $\mathbf{1} \to A$ is called a *point* of A; it is also called a *global element*. In general, an object A might have no points: we shall see examples in $\mathcal{E}$ and $\mathcal{G}$ in a moment—and more importantly for us, the Kochen-Specker theorem is the statement that a certain object has no points (cf. Section 4).

I apply these definitions to my examples of categories. In $\mathcal{S}$, each singleton set is a terminal object; we fix on one, often written $\{*\}$, as $\mathbf{1}$. (Dually, the empty set $\emptyset$ is initial.) Every non-empty set A has points, and indeed the points of A give a 'listing' of the elements of A. Furthermore, $\mathbf{1}$ *separates* arbitrary arrows (i.e. functions) in the sense that: given two functions $f : A \to B$ and $g : A \to B$, if for every point $x : \mathbf{1} \to A$ of A, we have $f \circ x = g \circ x$, then $f = g$.

In $\mathcal{E}$, each singleton set $\{*\}$ equipped with the identity endomap $* \mapsto *$ is a terminal object; i.e. $(\{*\}; id_{\{*\}})$ is a terminal object. I will call this terminal object, unique upto isomorphism, 'the loop'. Then a point of an object $(X; \alpha)$ is an arrow from the loop; and so corresponds to a fixed point of $(X; \alpha)$, i.e. an $x \in X$ such that $\alpha(x) = x$. So an object $A = (X; \alpha)$ need not have any points, and if it does, the points of the object do not in general give a 'listing' of all its 'elements' (unlike the situation in $\mathcal{S}$).

In $\mathcal{G}$ the situation is similar to $\mathcal{E}$. Here, a graph with just one dot and one arc (from the dot as source to the dot as target) is terminal; we can write it as $(s, t : \{*\} \to \{*\})$, and I again call it 'the loop'. Then a point of an arbitrary graph $(s, t : R \to D)$ corresponds to any loop in the graph, i.e. an arc $r \in R$, and a dot $d \in D$ such that $s(r) = d = t(r)$. So as in $\mathcal{E}$, an object in $\mathcal{G}$ need not have any points, and if it does, the points do not in general give a 'listing' of all its 'elements'.

The idea that in general, the points of an object do not give a 'listing of all its elements' can be made more precise in terms of the idea, mentioned above for $\mathcal{S}$, of the terminal object separating arbitrary arrows. Explaining this will motivate the idea of an generalized element.

In general, terminal objects do not separate arbitrary arrows in the way that in $\mathcal{S}$, $\mathbf{1}$ separates arbitrary functions, i.e. by a 'cancelling on the right' ('cancelling when done first') for all points x. For example, in $\mathcal{E}$ (resp. $\mathcal{G}$), the terminal object $\mathbf{1}$, i.e. the loop considered as a set-with-endomap (resp. considered as a graph), does not separate arbitrary arrows in $\mathcal{E}$ (resp. $\mathcal{G}$). The reason is that roughly speaking, points 'are' loops, and a set-with-endomap (resp. graph) can have plenty of structure apart from its loops, so that distinct arrows f and g can 'agree on all loops x' in the sense that for all $x, f \circ x = g \circ x$.

On the other hand, in many categories (including our examples, $\mathcal{E}$ and $\mathcal{G}$), there are one or two objects that *do* perform this separation of arbitrary arrows. In $\mathcal{E}$, the object in question 'is' the natural numbers with the successor map, i.e. $N := (Z^+, \sigma)$ with $Z^+ = \{1, 2, 3, ...\}$ and $\sigma(n) = n + 1$. That is to say, for any objects A, B in $\mathcal{E}$, and arrows $f : A \to B$ and $g : A \to B$: if for all arrows $x : N \to A$, we have $f \circ x = g \circ x$, then $f = g$. (Reason: the various x-images of N are all possible 'chains' in A, and f and g agreeing on their images of all possible chains implies that $f = g$).

In $\mathcal{G}$, the separation is done by a pair of (simple but not terminal) objects: the 'generic arc', Arc say, defined as having two dots and one arc from one to the other, and the 'naked dot', Dot say, defined as having one dot and no arcs. That is to say: for any objects (i.e. graphs) A, B in $\mathcal{G}$, and arrows $f : A \to B$ and $g : A \to B$: if (i) for all arrows $x : Arc \to A$, we have $f \circ x = g \circ x$ and also (ii) for all arrows $x : Dot \to A$, we have $f \circ x = g \circ x$; then $f = g$. (Reason: the various x-images of Arc and Dot code all the structure in A. For the component of an arrow $x : Arc \to A$ that describes x's action on dots, i.e. the 'dot-function' f_{Do}, can map two dots to one, so that Arc can be mapped into a loop in A.)

To sum up: the way that the various images of these special 'separating' objects—N (resp. Arc and Dot)—give a 'listing of the structure' of an arbitrary object A in the category $\mathcal{E}$ (resp. $\mathcal{G}$), generalizes the idea that in $\mathcal{S}$ the points of A give a 'listing' of all A's elements.

This suggests that we can think of an image of N (resp. Arc and Dot) as a kind of generalized (because structured) element of an arbitrary object A in $\mathcal{E}$ (resp. $\mathcal{G}$). And to avoid the vagueness of the word 'image', we can instead think of the *arrow itself*, i.e. $f : N \to A$ in $\mathcal{E}$ (resp. the arrows $f : Arc \to A$ and $g : Dot \to A$ in $\mathcal{G}$) as a generalized element of A.

We now generalize this idea to domains other than the special 'separating' objects, N (resp. Arc and Dot). That is to say, for any category we say that any arrow $f : B \to A$ is a *generalized element* of A. We also say f is a generalized element of A defined over B, or is a B-*element* of A, and we write $f \in_B A$. More vividly, we say B is the *stage of definition* of f, and that f is a *figure of shape* B, or a B-*figure*, in A. We often use the notation $f(x)$ in place of $f \circ x$ when x is a generalized element of the domain of f. It follows that in any category, the collection of all generalized elements of A separates arbitrary arrows to A in the above sense: given $f : B \to A$ and $g : B \to A$, if for all stages of definition C and C-elements of B (i.e. for all $x : C \to B$) we have $f \circ x = g \circ x$, then $f = g$.

The idea of a generalized element $f : B \to A$ of A, and its special case where f is monic, will be important in the definition of a subobject classifier.

3 Toposes and Partial Truth

I turn to developing one clause of the definition of a topos: the requirement that a topos contains a subobject classifier Ω. So far I have introduced Ω as a collection of generalized truth-values, generalizing the usual set $\{0, 1\}$. But from now on, the leading idea will be that Ω (together with some associated mathematical structure) generalizes the way that in set-theory characteristic functions classify whether an element x is in a given subset K of a set X. (This emphasis reflects Section 1's strategic decision to emphasise the general idea of partial truth (and so of parts), rather than logico-algebraic aspects of toposes.)

Thus we recall that for any set X, and any subset $K \subseteq X$, there is a characteristic function $\chi_K : X \to \{0, 1\}$, with $\chi_K(x) = 1$ or 0 according as $x \in K$ or $x \notin K$. Thus χ_K classifies the various x for the set-theoretically natural question (with only two possible answers!), "$x \in K$?". Furthermore, the structure of $\mathcal{S}$ secures the existence of the set of truth-values and the various functions χ_K: $\{0, 1\}$ is itself a set, i.e. an object in $\mathcal{S}$, and for each K, X with $K \subseteq X$, χ_K is an arrow from X to $\{0, 1\}$.

The main idea of the subobject classifier—as the name suggests—is that it is possible to formulate this 'classifying action' of the various χ_K in general category-theoretic terms, so as to give a fruitful generalization. I shall first generalize the idea of a subset of a given set, to the idea of a subobject of an object in an arbitrary category $\mathcal{C}$; and explain the notion of a pullback (Section 3.1). Then I shall give the definition of a subobject classifier (Section 3.2); and illustrate the definition in the toposes $\mathcal{E}$ and $\mathcal{G}$ (Section 3.3), and finally in toposes of presheaves (Section 3.4).

3.1 Monics as Parts, Pullbacks

In any category, one can define a categorial analogue of the set-theoretic idea of subset: it is called a 'subobject'. This notion is not only important for explaining subobject classifiers. It also provides a formal generalization of the notion of part as studied in mereology, so as to allow for 'structured parts'. I shall indicate this line of thought in this Subsection, but not pursue it in detail.

One starts from the idea that the subsets of a set X are in one-to-one correspondence with certain injective functions to X: a subset K of X corresponds to the injective (*i.e.*, one-to-one) function $K \to X$ sending $x \in K$ to $x \in X$. Since category theory provides a generalization of injective functions, viz. the monic arrows (for short, monics), in any category one defines a *subobject* of any object X to be: a monic with codomain X. One also writes a monic with a 'hook-arrow', $k : K \hookrightarrow X$.

At first sight, this definition seems to be formally correct, but to have an informal defect; which I shall briefly discuss in turn. We immediately see that the definition is formally right, at least for the categories we have discussed, $\mathcal{E}$ and $\mathcal{G}$. For we saw in Section 2.2's discussion of monics in these categories, that if $k : K \hookrightarrow X$ is a monic (in $\mathcal{E}$ or in $\mathcal{G}$) then the 'copy' of K inside X (the 'range' of the arrow k) 'looks exactly like' K itself. So for these categories at least, the definition of a subobject as a monic captures a notion of 'structured part'.

But on the other hand, this definition of subobject clearly ignores the underlying identity of the 'elements' that the objects in the category are comprised of. And it does so in $\mathcal{S}$, no less than in other categories such as $\mathcal{E}$ and $\mathcal{G}$. For a subset K of X in $\mathcal{S}$, it is the very same element x that is in K and in X: that is why it is natural to call K a part of X. But there is no such requirement for a monic: x in K can be sent by a monic k to a very different y in X. (Besides, K is not itself a monic and so *not* a subobject, according to this definition.) Of course, ignoring the underlying identity of basic elements is well-nigh universal in mathematics (and a springboard to structuralism in the philosophy of mathematics); and is all the more natural in category theory. But from the philosophical perspective of trying to generalize mereology or the notion of part, it seems a defect.

But it turns out that, despite this defect, the definition of a subobject as a monic does give a successful generalization of several formal aspects of mereology, such as the subobjects of an object having an inclusion relation between them. It is worth briefly indicating how this generalization proceeds, partly because it gives an example of the usefulness of Section 2.3's notion of a generalized element.

Given a monic $i : B \hookrightarrow A$, we say that a C-element $x \in_C A$, i.e. an arrow $x : C \to A$, is a *member* of i, and write $x \in i$, if x factors through i; that is, if there is some $h : C \to B$ such that $i \circ h = x$. This definition is natural (again, *modulo* the 'defect' above): it is clear that in $\mathcal{S}$, $x \in i$ iff $\mathrm{ran}(x) \subseteq \mathrm{ran}(i)$, and (exercise!) similar points hold in in $\mathcal{E}, \mathcal{G}$.

We now define an inclusion relation on monics to any one object. Given monics $i : A \hookrightarrow C$ and $j : B \hookrightarrow C$, we write $i \subseteq j$, and say that i is *included* in j, if i factors through j: that is, if there is an arrow $s : A \to B$ such that $j \circ s = i$. Again this definition is natural. It gives the intuitively right results in $\mathcal{S}, \mathcal{E}, \mathcal{G}$. And in any category, there are two nice elementary results: (1) Since j is monic, if there is any such s there is only one; and it is itself monic. So if i is included in j (i.e. $i \subseteq j$), it is included in only one way, and the inclusion s is itself a sub-object of the domain of j. (2) Given two subobjects of C, $i : A \hookrightarrow C$ and $j : B \hookrightarrow C$, $i \subseteq j$ iff every member of i (in the sense of the previous paragraph) is a member of j.

We then define: i is *equivalent* to j, $i \equiv j$, iff both $i \subseteq j$ and $j \subseteq i$. Again, this is mathematically natural. From $i \equiv j$ it follows that the domains of i and j are isomorphic; and result (2) implies that $i \equiv j$ iff i and j have exactly the same members. Besides, we can in general now recover the usual notion of a part (the notion respecting the 'underlying identity' of objects' elements—in the usual sense of 'element'!), by picking within an equivalence class of monics, the appropriate representative element (or rather its domain). We will see an example of this in Section 3.4's discussion of subobjects of presheaves.

Here is a final example of how generalized elements allow us to generalize mereological ideas. We can make precise the idea of 'that part of A on which a pair of parallel arrows $f : A \to B$ and $g : A \to B$ agree', as follows. Given a pair of parallel arrows, $f : A \to B$ and $g : A \to B$, we say that a sub-object $e : E \hookrightarrow A$ of A is an *equalizer* for f and g iff: for every C and every $x \in_C A$, x is a member of e iff $f \circ x = g \circ x$. (Or as we wrote more intuitively at the end of Section 2.3: iff $f(x) = g(x)$).

So much by way of formal generalization of the notion of part. My discussion of the subobject classifier needs one more preliminary notion: the notion of a pullback. We define a *pullback* of two arrows with common codomain, $f : A \to C$ and $g : B \to C$ to be an object P and arrows $p_1 : P \to A$ and $p_2 : P \to B$, such that:
(i) the 'square of arrows' is a commutative diagram in the usual sense, i.e. $f \circ p_1 = g \circ p_2$; and
(ii) for any T and any $h : T \to A$ and $k : T \to B$ such that the 'outer square' commutes, i.e. $f \circ h = g \circ k$, there is a unique $u : T \to P$ such that the entire diagram commutes, i.e. $p_1 \circ u = h$ and $p_2 \circ u = k$.
The maps p_1 and p_2 are called *projections*. Pullbacks are defined upto unique isomorphisms, which multiply one pair of projections to give the other.

Unfortunately, the best intuitive explanation of this rather abstract definition uses the notion of a product, $A \times B$ of two objects A and B; a notion I have not taken space to define. But here it suffices to make two comments. (1) In Section 3.2, the abstract definition will be enough for us. (2) (For readers who know the notion of product, i.e. the category-theoretic generalization of set-theory's cartesian product): If the product $A \times B$ exists, then a pullback P is a subobject of the product $A \times B$, in fact an equalizer. More precisely: suppose we are given $f : A \to C$ and $g : B \to C$, and that the product $A \times B$ exists, with projections (for the product) $p_1 : A \times B \to A, p_2 : A \times B \to B$. Then: (i) If there is an equalizer for the arrows, $f \circ p_1$ and $g \circ p_2$, then this equalizer is a pullback (with its projections being p_1, p_2). (ii) Conversely, if there is a pullback for f and g, then it has an arrow to $A \times B$ and that arrow is an equalizer.

3.2 The Subobject Classifier

Recall the main idea of a subobject classifier, Ω. It is an object in the topos, for which—just as for subsets of a given set X and characteristic functions in $\mathcal{S}$—there is to be a one-to-one correspondence between subobjects of any given object X, and arrows from X to Ω.

Formally, we define: A *subobject-classifier* is an object Ω and a point (global element) $t : \mathbf{1} \to \Omega$ such that for any monic $k : K \hookrightarrow X$ there is a unique $\chi_k : X \to \Omega$ that makes the following square:

$$\begin{array}{ccc} K & \xrightarrow{!_K} & \mathbf{1} \\ \downarrow{\scriptstyle k} & & \downarrow{\scriptstyle t} \\ X & \xrightarrow{\chi_k} & \Omega \end{array} \tag{3.1}$$

into a pullback; where we have written $!_K$ for the unique arrow from K to $\mathbf{1}$. That is to say, we require:
(i) this square must commute, i.e. $K \stackrel{!_K}{\to} 1 \stackrel{t}{\to} \Omega$ and $K \stackrel{k}{\hookrightarrow} X \stackrel{\chi_k}{\to} \Omega$ are the same arrow;
(ii) for any T and any $h : T \to 1$ and $x : T \to X$ such that the 'outer square' commutes, ie $t \circ h = \chi_k \circ x$, there is a unique $u : T \to K$ such that the entire diagram commutes, i.e. $!_K \circ u = h$ and $k \circ u = x$.

The intuitive idea of this definition is readily expressed for the case where objects of the topos are structured sets, as follows.

- (1) Think of Ω as a (structured) set of truth-values; think of $t : 1 \to \Omega$ as picking out a designated element, call it *true*, of Ω (intuitively, 'totally true').
- (2) Writing the unique arrow from any object T to $\mathbf{1}$ as $!_T : T \to \mathbf{1}$, (1) means that for any T, the composite $t \circ !_T : T \to \Omega$ takes all of T to *true*. In particular, this is true for the case $T := K$.
- (3) The fact that the square commutes then implies that χ_k takes all the elements of X that are 'in K', i.e. elements of X that are 'in the copy of K contained in X as the range of k', into the designated value *true* within Ω.
- (4) The requirement that the square be a pullback relates to generalized elements of X. That is, since these generalized elements are just arrows $x : T \to X$, for arbitrary objects T, the requirement of being a pullback makes χ_k have the correct action on these elements. To be precise; first recall from Section 3.1 that we say a T-element x of X is a member of k, and write $x \in k$ iff x factors through k. And when the monic k is clear from the context, there is a common abuse of notation: we say that x is a member of the domain K, rather than the monic k, and we write $x \in K$, rather than $x \in k$.

  Then the point is that our definition implies: the classifying arrow χ_k of any $k : K \hookrightarrow X$ is the unique arrow such that for any T-element x of X, x is a member of K [properly speaking, of k] iff $\chi_k(x) := \chi_k \circ x = t \circ !_T$. In short, $x \in K$ iff $\chi_k(x) = true$.

To sum up (1) to (4): the definition of the subobject classifier makes χ_k have the correct classifying action on elements of X, even on generalized elements of X, in the sense that for any such element x, $x \in K$ iff $\chi_k(x) = true$. (Besides, the definition requires that χ_k be the unique arrow with this correct action.)

I should stress three other general features of subobject classifiers, before turning to examples. These features are not obvious from the definition I have given, but follow from it together with the other clauses in the definition of a topos. (A) Not only does Ω with its arrows χ_k correctly classify elements of X, as regards whether they are 'in K'. It also correctly classifies them, as regards whether they are 'partly in K'. That is, an element x of X, even a generalized element $x : T \to X$, gets sent by χ_k to the partial truth-value in Ω (a truth-value other than *true*) that correstly classifies the 'extent' to which x is in K. This will become clearer in Sections 3.3 and 3.4's examples.

(B) Ω is fixed by the structure of the topos concerned in the precise sense that, although the above definition of Ω characterizes Ω solely in terms of conditions on the topos' objects and arrows, Ω is provably unique (up to isomorphism).

(C) In any topos, Ω has a natural logical structure. More exactly, Ω has the internal structure of a Heyting algebra object: the algebraic structure appropriate for intuitionistic logic, mentioned in Section 1. Furthermore, in any topos, the collection of subobjects of any given object X is a complete Heyting algebra (a locale), written Ω^X. I shall not discuss this Heyting algebra structure in detail, but note only that its presence confirms my remarks 1., 3. and 4. in Section 1.

3.3 Ω in $\mathcal{E}$ and $\mathcal{G}$

I turn to displaying Ω in $\mathcal{E}$ and $\mathcal{G}$. As we shall see, the generalized truth-values (intuitively, the 'elements' of Ω) are the answers to a natural multiple-choice question about the objects in the topos, just as "$x \in K$?" is natural for sets.

In $\mathcal{E}$: Recall that a set X equipped with a given function $\alpha : X \to X$ is called an *endomap*, written $(X; \alpha)$; and the family of all endomaps forms a category—indeed, a topos—when one defines an arrow from $(X; \alpha)$ to $(Y; \beta)$ to be an ordinary set-function f between the underlying sets, from X to Y, that preserves the endomap structure, i.e. $f \circ \alpha = \beta \circ f$.

Applying the definition of a subobject, it turns out that a subobject of $(X; \alpha)$ is a subset of X that is closed under α, equipped with the restriction of α: i.e., a subobject is $(K, \alpha \mid_K)$, with $K \subseteq X$ and such that $\alpha(K) \subset K$. So a natural question, given $x \in X$ and a subendomap $(K, \alpha \mid_K)$, is: "How many iterations of α are needed to send x (or rather its descendant, $\alpha(x)$ or $\alpha^2(x)$ or $\alpha^3(x)$ ...) into K?" The possible answers are '0 (*i.e.*, $x \in K$)', '1', '2',..., and 'infinity (i.e., the descendants never enter K)'; and if the answer for x is some natural number n (resp. 0, infinity), then the answer for $\alpha(x)$ is $n - 1$ (resp. 0, infinity). So the possible answers can be presented in a diagram, as in Figure 1:

And it turns out that this is exactly the object Ω in the category of endomaps! To sum up this example: Ω is the collection of possible answers to "At what stage do you (strictly speaking: your descendant under α) enter K?" This notion of a *stage* will come to the fore in Sections 3.4 and 4.

I turn to $\mathcal{G}$. Recall that an object in $\mathcal{G}$ is a graph; we can think of a graph as a set of dots equipped with a (not necessarily symmetric) binary relation; where we think of the binary relation as a set of arcs among the dots. A subobject of a graph 'is' a subgraph; (more exactly, a corresponding monic with the given subgraph as 'range'). And we can think of a subgraph as a subset of the given graph's set of dots, equipped with a subset of the given graph's set of arcs.

So a natural question about the dots and arcs of a given graph X, in relation to a subgraph K of X, is: "Are you in K or not?". For a dot, there are only two possibilities: a dot in X is simply either in, or out, of a subgraph K. But for an arc in X, there are five possible relations to K!
(i) The arc can be wholly in K, in the sense that not only are the arc's source dot and target dot in K, but so is the arc itself.
(ii) The arc can be 'just out' of K in the sense that it is out, but its source and target dots are in.
(iii) The arc is out, and so is its source dot; but its target dot is in.
(iv) The arc is out, and so is its target dot; but its source dot is in.
(v) The arc can be 'wholly out' of K in that it is out, and so are its source and target dots.
These possibilities—two for dots, and five for arcs—can be pictured in a graph. And indeed it

turns out that Ω is exactly this graph (upto isomorphism)! Figure 2 shows an example of how this Ω classifies the dots and arcs of a graph, as to whether they are in the subgraph K indicated by the dashed circle: the arcs of the given graph have Roman letters, and the arcs in Ω that are 'their truth-values' have the corresponding Greek letters.

To sum up this example: Ω is the collection of possible answers to "What part of you is in K?" (where the source and target dots of a given arc count as part of it—but a dot has no proper parts!). This notion of a *part* of an element x being in a subobject will arise again in Section 4.

3.4 Toposes of Presheaves

As discussed in Section 1, our proposal about partial truth in a physical theory uses a topos of presheaves. So to prepare for Section 4's discussion of that proposal, I now introduce presheaves.

We first need the idea of a covariant functor between a pair of categories $\mathcal{C}$ and $\mathcal{D}$. Intuitively, this is a arrow-preserving function from one category to the other. To be precise: A *covariant functor* $\mathbf{F}$ from a category $\mathcal{C}$ to a category $\mathcal{D}$ is a function that assigns

1. to each $\mathcal{C}$-object A, a $\mathcal{D}$-object $\mathbf{F}(A)$;

2. to each $\mathcal{C}$-arrow $f : A \to B$, a $\mathcal{D}$-arrow $\mathbf{F}(f) : \mathbf{F}(A) \to \mathbf{F}(B)$ such that: (i) $\mathbf{F}(\mathrm{id}_A) = \mathrm{id}_{\mathbf{F}(A)}$; and (ii) if $g : C \to A$, and $f : A \to B$ then $\mathbf{F}(f \circ g) = \mathbf{F}(f) \circ \mathbf{F}(g)$.

A *presheaf* on the category $\mathcal{C}$ is defined to be a covariant functor $\mathbf{X}$ from $\mathcal{C}$ to $\mathcal{S}$, the category of sets; (and $\mathcal{C}$ is then called the *base-category*). A presheaf is also called a *varying set*, since the membership of the assigned set $\mathbf{X}(A)$ varies from one object A in $\mathcal{C}$ to another.

Note that given a $\mathcal{C}$-arrow $f : A \to B$: if $\mathbf{X}(A) \neq \emptyset$, then $\mathbf{X}(B) \neq \emptyset$. So as we go 'downstream', the assigned set $\mathbf{X}(A)$ cannot 'disappear', becoming $\emptyset$: "once in, the descendants are forever in". But $\mathbf{X}$ can 'shrink' in that $\mathbf{X}(f)$ can be non-injective; and $\mathbf{X}$ can of course 'grow' (viz. if the codomain of $\mathbf{X}(f)$ is a superset of its range).

To make the collection of all presheaves on $\mathcal{C}$ into a category, we need to define what is meant by an arrow between two presheaves $\mathbf{X}$ and $\mathbf{Y}$. The intuitive idea is that an arrow from $\mathbf{X}$ to $\mathbf{Y}$ gives a 'picture' of $\mathbf{X}$ within $\mathbf{Y}$. Formally, we define an arrow to be a *natural transformation* $N : \mathbf{X} \to \mathbf{Y}$, by which is meant a family of functions (called the *components* of N) $N_A : \mathbf{X}(A) \to \mathbf{Y}(A)$, A an object in $\mathcal{C}$, such that if $f : A \to B$ is an arrow in $\mathcal{C}$, then the composite function $\mathbf{X}(A) \xrightarrow{N_A} \mathbf{Y}(A) \xrightarrow{\mathbf{Y}(f)} \mathbf{Y}(B)$ is equal to $\mathbf{X}(A) \xrightarrow{\mathbf{X}(f)} \mathbf{X}(B) \xrightarrow{N_B} \mathbf{Y}(B)$. In other words, we have the commutative diagram

$$\begin{array}{ccc} \mathbf{X}(A) & \xrightarrow{\mathbf{X}(f)} & \mathbf{X}(B) \\ \downarrow{\scriptstyle N_A} & & \downarrow{\scriptstyle N_B} \\ \mathbf{Y}(A) & \xrightarrow{\mathbf{Y}(f)} & \mathbf{Y}(B) \end{array} \tag{3.2}$$

The resulting category of presheaves on $\mathcal{C}$ is denoted $\mathcal{S}^{\mathcal{C}}$.

As in Section 3.1, a subobject of $\mathbf{X}$ is formally defined as a monic arrow (i.e. monic natural transformation) with codomain $\mathbf{X}$. But recalling our discussion of equivalent subobjects, we see that each equivalence class of subobjects of $\mathbf{X}$ has a member $i : \mathbf{K} \hookrightarrow \mathbf{X}$ that 'respects the underlying identity of elements': that is, i is such that for each A, the component map

$i_A : \mathbf{K}(A) \to \mathbf{X}(A)$ is a subset embedding, *i.e.*, $\mathbf{K}(A) \subseteq \mathbf{X}(A)$. Thus, if $f : A \to B$ is any arrow in $\mathcal{C}$, we get the analogue of the commutative diagram Eq. (3.2):

$$
\begin{array}{ccc}
\mathbf{K}(A) & \xrightarrow{\mathbf{K}(f)} & \mathbf{K}(B) \\
\Big\downarrow{\scriptstyle i_A} & & \Big\downarrow{\scriptstyle i_B} \\
\mathbf{X}(A) & \xrightarrow{\mathbf{X}(f)} & \mathbf{X}(B)
\end{array}
\tag{3.3}
$$

but where now the vertical arrows are subset inclusions.

From now on, we fix on such 'identity-respecting' subobjects, for simplicity; (but without loss of generality, since every equivalence class of subobjects contains such). Then $\mathbf{K}(f)$ is the restriction of $\mathbf{X}(f)$ to $\mathbf{K}(A)$; and this implies remarks about "once in $\mathbf{K}$, descendants are forever in", and about $\mathbf{K}$ 'shrinking' or 'growing' as we go 'downstream', corresponding to the remarks above about $\mathbf{X}$. Thus the fact that $\mathbf{K}(f) = \mathbf{X}(f) \mid_{\mathbf{K}(A)}$ implies that for $x, y \in \mathbf{K}(A), \mathbf{K}(f)(x) = \mathbf{K}(f)(y)$ only if $\mathbf{X}(f)(x) = \mathbf{X}(f)(y)$. So $\mathbf{K}$ can only 'shrink' as we move from $\mathbf{K}(A)$ to $\mathbf{K}(B)$ to the extent allowed by $\mathbf{X}$ 'shrinking'. But $\mathbf{K}$ can 'grow' from $\mathbf{K}(A)$ to $\mathbf{K}(B)$, even if $\mathbf{X}$ does not; for $\mathbf{K}(B)$ can be a proper superset of ran$(\mathbf{K}(f))$. This discussion of subobjects in $\mathcal{S}^{\mathcal{C}}$ is summed up in Figure 3.

It turns out that the category $\mathcal{S}^{\mathcal{C}}$ is a topos. So for us, the question is: what is the structure of its subobject classifier $\mathbf{\Omega}$? (I write $\mathbf{\Omega}$, not Ω, since it is itself a preheaf.) It turns out that, as with our previous examples, $\mathbf{\Omega}$ encodes the possible answers to a natural multiple-choice question one can ask of the 'elements' of which objects in the topos are built. For presheaves, the question is asked of an element $x \in \mathbf{X}(A)$ at a 'stage' A in $\mathcal{C}$: "At which stages B in $\mathcal{C}$ are the 'descendants' of x, i.e. $\mathbf{K}(f)(x) = \mathbf{X}(f)(x)$, elements of $\mathbf{K}(B)$?" Note incidentally, that the topic, 'when do x's descendants under a certain function enter a certain set?', is reminiscent of Ω for the category of endomaps $\mathcal{E}$.

Given this question, and recalling that in a topos a subobject $\mathbf{K}$ of an object $\mathbf{X}$ corresponds to a characteristic arrow, call it $\chi^K : \mathbf{X} \to \mathbf{\Omega}$, we intuitively expect that an element $x \in \mathbf{X}(A)$ will be mapped by χ^K to the set of those B, with $f : A \to B$, at which $\mathbf{X}(f)(x) \in \mathbf{K}(B)$. So we expect the various partial truth-values to be various sets of objects in the base-category $\mathcal{C}$.

This expectation is essentially correct, but needs two qualifications. (1) Since χ^K is an arrow in $\mathcal{S}^{\mathcal{C}}$, i.e. a natural transformation, it will consist of a component function $\chi^{\mathbf{K}}_A$ at each A in $\mathcal{C}$. So the objects A in the base-category $\mathcal{C}$ become *stages, contexts*, relative to which truth-values are ascribed; (in addition to their role as stages at which 'descendants' enter the set $\mathbf{K}(A)$).

(2) We need to allow for the possibility (not shown in Fig. 3) that there are two $\mathcal{C}$-arrows f, g from A to B (or even more than two). This means there can be distinct functions $\mathbf{X}(f)$ and $\mathbf{X}(g)$ from $\mathbf{X}(A)$ to $\mathbf{X}(B)$; and so for a given $x \in \mathbf{X}(A)$, we can have $\mathbf{X}(f)(x) \in \mathbf{K}(B)$ but $\mathbf{X}(g)(x) \notin \mathbf{K}(B)$. The most straightforward way that $\mathbf{\Omega}$ could encode these possibilities—i.e. assign different truth-values according to which of the functions $\mathbf{X}(f)$ and $\mathbf{X}(g)$ send x into $\mathbf{K}(B)$—is for the truth-values to contain the $\mathcal{C}$-arrows themselves, f etc., rather than their codomains B etc. And indeed, this is exactly what $\mathbf{\Omega}$ does.

Besides, the feature that, going 'downstream' from A, once a descendant $\mathbf{X}(f)(x)$ of x gets into a subobject $\mathbf{K}$ of $\mathbf{X}$, all subsequent descendants are also in $\mathbf{K}$, implies that a set S of arrows that is a truth-value at A—i.e. a set of arrows with domain A, at whose codomain B $\mathbf{X}(f)(x) \in \mathbf{K}(B)$—is 'closed under going downstream' in the sense that: if $f : A \to B$ belongs to S, and if $g : B \to C$ is any arrow, then $g \circ f : A \to C$ also belongs to S. A set of arrows each with domain A that has this property is called a *sieve* on A. So we expect the sieves on A to be the generalized truth-values at A; and since the subobject classifier $\mathbf{\Omega}$ is to be a presheaf on $\mathcal{C}$, we expect $\mathbf{\Omega}$'s assignment to each A in $\mathcal{C}$ to be the set of all sieves on A.

That leaves open how to define $\mathbf{\Omega}$'s action on $\mathcal{C}$-arrows f. But it turns out that a simple 'push-along' action, as in the following precise definition of $\mathbf{\Omega}$, works. The presheaf $\mathbf{\Omega} : \mathcal{C} \to \mathrm{Set}$ is defined by:

(i) For A in $\mathcal{C}$, $\mathbf{\Omega}(A)$ is to be the set of all sieves on A;

(ii) if $f : A \to B$, then $\mathbf{\Omega}(f) : \mathbf{\Omega}(A) \to \mathbf{\Omega}(B)$ is defined for all $S \in \mathbf{\Omega}(A)$ as

$$\mathbf{\Omega}(f)(S) := \{h : B \to C \mid h \circ f \in S\}. \tag{3.4}$$

We can now spell out how $\mathbf{\Omega}$ works as the subobject classifier for $\mathcal{S}^{\mathcal{C}}$. Let $\mathbf{K}$ be a subobject of $\mathbf{X}$. Then there is an associated *characteristic* arrow $\chi^{\mathbf{K}} : \mathbf{X} \to \mathbf{\Omega}$, whose 'component' $\chi^{\mathbf{K}}_A : \mathbf{X}(A) \to \mathbf{\Omega}(A)$ at each 'stage of truth' A in $\mathcal{C}$ is defined as

$$\chi^{\mathbf{K}}_A(x) := \{f : A \to B \mid \mathbf{X}(f)(x) \in \mathbf{K}(B)\} \tag{3.5}$$

for all $x \in \mathbf{X}(A)$. That the right hand side of Eq. (3.5) actually *is* a sieve on A follows from the definition of a subobject. Thus, in each 'branch' of $\mathcal{C}$ going 'downstream' from the stage A, $\chi^{\mathbf{K}}_A(x)$ picks out the first member B in that branch for which $\mathbf{X}(f)(x)$ lies in the subset $\mathbf{K}(B)$, and the commutative diagram Eq. (3.3) then guarantees that $\mathbf{X}(h \circ f)(x)$ will lie in $\mathbf{K}(C)$ for all $h : B \to C$. Thus each 'stage of truth' A in $\mathcal{C}$ serves as a possible context for an assignment to each $x \in \mathbf{X}(A)$ of a generalised truth-value.

There is a converse to Eq. (3.5): namely, each arrow $\chi : \mathbf{X} \to \mathbf{\Omega}$ (*i.e.*, a natural transformation between the presheaves $\mathbf{X}$ and $\mathbf{\Omega}$) defines a subobject $\mathbf{K}^{\chi}$ of $\mathbf{X}$ via

$$\mathbf{K}^{\chi}(A) := \chi_A^{-1}\{1_{\mathbf{\Omega}(A)}\}. \tag{3.6}$$

at each stage of truth A. To sum up Eq. (3.5) and (3.6): we have a one-to-one correspondence between the subobjects of $\mathbf{X}$ and arrows $\chi : \mathbf{X} \to \mathbf{\Omega}$.

Two final remarks on $\mathbf{\Omega}$. (1) First, a logico-algebraic remark. The set $\mathbf{\Omega}(A)$ of sieves on any A in $\mathcal{C}$ has the structure of a Heyting algebra, under natural operations. I shall not define 'Heyting algebra' (it is a kind of distributive lattice), but only indicate how natural the logical operations in $\mathbf{\Omega}(A)$ are. The partial order is defined by $S_1 \leq S_2$ iff $S_1 \subseteq S_2$; so that the 1 of the algebra (the 'totally true' truth-value, *true*) is the set of all arrows with domain A (called the *principal sieve* on A), and the 0 of the algebra is the empty set of sieves. The logical connectives, 'and', 'or' are defined by:

$$S_1 \wedge S_2 := S_1 \cap S_2 \tag{3.7}$$

$$S_1 \vee S_2 := S_1 \cup S_2 \tag{3.8}$$

(2) I turn to points (global elements) of a presheaf. In $\mathcal{S}^{\mathcal{C}}$, a terminal object $\mathbf{1} : \mathcal{C} \to \mathcal{S}$ can be defined by: (i) $\mathbf{1}(A) := \{*\}$ at all stages A in $\mathcal{C}$; (ii) if $f : A \to B$ is an arrow in $\mathcal{C}$ then $\mathbf{1}(f) : \{*\} \to \{*\}$ is defined to be the function $* \mapsto *$. This is indeed a terminal object since, for any presheaf $\mathbf{X}$, there is a unique natural transformation $N : \mathbf{X} \to \mathbf{1}$ whose components $N_A : \mathbf{X}(A) \to \mathbf{1}(A) = \{*\}$ are the constant maps $x \mapsto *$ for all $x \in \mathbf{X}(A)$.

So a point (global element, also called *global section*) of a presheaf $\mathbf{X}$, i.e. an arrow $\gamma : \mathbf{1} \to \mathbf{X}$, corresponds to a choice of an element $\gamma_A \in \mathbf{X}(A)$ for each stage A in $\mathcal{C}$, such that, if $f : A \to B$, the 'matching condition'

$$\mathbf{X}(f)(\gamma_A) = \gamma_B \tag{3.9}$$

is satisfied. Some presheaves have no points: a 'globally matching' choice of element cannot be made. Examples include certain presheaves that arises naturally in any quantum theory: and the non-existence of such points is exactly the content of the Kochen-Specker theorem.

4 Presheaves of Propositions and Sieve-Valued Valuations in Physics

Finally, I sketch our proposal applying the material in Section 3 to physical theories. As discussed in Section 1, this application is laid out in detail elsewhere; so here I will be very brief, emphasising only the connections with this paper's ideas of partial truth.

The fundamental ideas of the proposal are:

1. To make a category, $\mathcal{O}$ say, out of a theory's physical quantities. Each quantity is to be an object in the category, and there is to be an $\mathcal{O}$-arrow $f : A \to B$ iff $B = f(A)$ in the usual sense of the functional calculus of quantities (i.e. roughly the value of B is f of the value of A).[4]

2. To define a *coarse-graining presheaf* $\mathbf{G}$ on $\mathcal{O}$ as follows:
(a) $\mathbf{G}$ assigns to each quantity A the set of propositions "$A \in \Delta$" (read as saying that the value of A lies in a Borel set of real numbers, $\Delta \subseteq \mathbb{R}$) for varying Δ; more exactly, $\mathbf{G}(A)$ is to be the set of these propositions' mathematical representatives, such as the algebra of A's spectral projectors on a quantum system's Hilbert space or the algebra of A's characteristic functions on a classical system's phase space; and
(b) $\mathbf{G}$ assigns to each $\mathcal{O}$-arrow $f : A \to B$ the 'coarse-graining function' $\mathbf{G}(f) : \mathbf{G}(A) \to \mathbf{G}(B)$ defined by $\mathbf{G}(f) : A \in \Delta \mapsto f(A) \in f(\Delta)$.

3. To assign as the truth-value of the proposition $A \in \Delta$, relative to the context A, a sieve on A, written $\nu(A \in \Delta)$. These sieve-valued valuations can be required to have various natural properties, among them a generalization for the presheaf $\mathbf{G}$ of the main premise of the Kochen-Specker theorem: viz. the $FUNC$ condition, that if $B = f(A)$, the real numbers r and r' that are the respective values of A and B are related by f, i.e. $r' = f(r)$.

4. To show that the states (both pure and mixed) of the physical theory (quantum or classical) define sieve-valued valuations enjoying such properties. I illustrate this with the case of a pure quantum state represented by a vector ψ in the Hilbert space $\mathcal{H}$. (The situation for a mixed quantum state is exactly analogous.) ψ defines a sieve on each bounded self-adjoint operator A on $\mathcal{H}$ (i.e. each A in $\mathcal{O}$) by the requirement that an $\mathcal{O}$-arrow $f : A \to B$ is in the sieve iff ψ is in the range of the spectral projector representing $f(A) \in f(\Delta)$ (i.e. the projector $\hat{E}[f(A) \in f(\Delta)]$, in a usual notation). That is, ψ defines a sieve by:

$$\begin{aligned} \nu^{\psi}(A \in \Delta) &:= \{f : A \to B \mid \hat{E}[B \in f(\Delta)]\psi = \psi\} \\ &= \{f : A \to B \mid \mathrm{Prob}(B \in f(\Delta); \psi) = 1\} \end{aligned} \tag{4.1}$$

where $\mathrm{Prob}(B \in f(\Delta); \psi)$ is the usual Born-rule probability that the result of a measurement of B will lie in $f(\Delta)$, given the state ψ.

This definition generalizes orthodox quantum theory's so-called eigenstate-eigenvalue link, in the sense that it requires, not that ψ be in the range of $\hat{E}[A \in \Delta]$, but only that ψ be in the range of the projector $\hat{E}[f(A) \in f(\Delta)]$; and this is a logically weaker requirement, since in the lattice $\mathcal{L}(\mathcal{H})$ of projectors on $\mathcal{H}$, $\hat{E}[A \in \Delta] \leq \hat{E}[f(A) \in f(\Delta)]$. Furthermore, one can check that the definition satisfies the presheaf version of $FUNC$, and also has other natural properties.

To sum up the proposal, as it applies to quantum theory: the new proposed truth-value of $A \in \Delta$ is to be determined by the set of quantities $f(A)$ for which the corresponding weaker proposition $f(A) \in f(\Delta)$ is true in the old (i.e., eigenstate-eigenvalue link) sense. To put it less exactly, but more memorably: the new proposed truth-value of a proposition is to be determined by the set of its consequences that are true in the old sense.[5]

From this summary, it is clear that this paper has added three motivations for this proposal. The first two are general. (1) I have given a general defence of partial truth as a notion that is made tractable and illuminating within a topos. (2) Section 3.4 motivated the definition of sieves, and thereby of $\mathbf{\Omega}$, in terms of a natural question one can ask about the elements x of an assigned set $\mathbf{X}(A)$: a question analogous to that encoded by the subobject classifier in $\mathcal{E}$.

(3) The third motivation relates to the last part of the summary above: the fact that the proposal defines the partial truth of $A \in \Delta$ in terms of which logically weaker propositions $f(A) \in f(\Delta)$ are wholly true. Since in philosophical logic it is natural (and common) to think of a proposition as the conjunction of its consequences, and to think of each consequence as part of the proposition, this feature of our proposal is intuitively appealing. For it implies that our proposal takes the partial truth of a proposition to be a matter of the whole truth of a part of the proposition. (And in this regard, the notion of partial truth in $\mathcal{S}^{\mathcal{O}}$ is like the notion of an arc being within a subgraph, within $\mathcal{G}$.)

To conclude: in this paper I have argued that topos theory provides a promising general framework for (i) making sense of partial truth, and (ii) developing a generalized mereology that considers a notion of 'structured part'. In particular, I have sketched how these notions apply to a recent proposal to use a topos of presheaves on a base-category of physical quantities, to define partial and contextual truth-values for propositions about the values of physical quantities in physical theories.[6]

5 Endnotes

1. Cf. Isham and Butterfield (1998, 2000), Butterfield and Isham (1999), Hamilton, Isham and Butterfield (2000).

2. Lawvere and Shanuel (1997) gives a very lucid and enjoyable introduction to categories and toposes, and discusses endomaps and graphs in detail.

3. But I will ignore the motivation for these notions in terms of determination and choice problems; i.e. the equivalence between an arrow having a retraction (resp. section) and determination (resp. choice) problems having a solution. For this, cf. e.g. Lawvere and Shanuel (1997).

4. Hamilton, Isham and Butterfield (2000) describe how the proposal carries over in all details to a more general notion of context, viz. as a commutative von Neumann algebra.

5. I should add that this summary of the proposal downplays some facts that are treated at length in the cited papers, including:
(i) The application to classical physical theories is perfectly analogous to the quantum theory

case. But it is the interpretative problems of quantum theory—specifically, the difficulties of a classical realism, as shown by no-go theorems like the Kochen-Specker theorem—that motivate partial truth-values for propositions "$A \in \Delta$".

(ii) The Kochen-Specker theorem is equivalent to the non-existence of points (global elements) of certain presheaves on $\mathcal{O}$ (for quantum theory). This equivalence illuminates the Kochen-Specker theorem, and its various proofs (cf. Hamilton 2000). But it is not essential to our positive proposal. In particular, $\mathbf{G}$ is not one of the presheaves that lack points.

(iii) The presheaf generalization of $FUNC$ enjoyed by the valuations ν is equivalent to ν specifying a subobject of $\mathbf{G}$; i.e. to ν's assignment at each quantity A, $A \in \Delta \mapsto \nu(A \in \Delta)$, defining a natural transformation from $\mathbf{G}$ to Ω.

6. It is a pleasure to thank: the conference organizers and editors for all their work; audiences in Cracow, Chicago, London and Salzburg; Chris Isham and John Hamilton for many discussions of this material—and for invaluable computing help!

References

Butterfield, J. and Isham, C: 1999, 'A topos perspective on the Kochen-Specker theorem: II. Conceptual aspects, and classical analogues', *International Journal of Theoretical Physics*, **38**, pp. 827–859. quant-ph/9808067.

Hamilton, J: 2000, 'An obstruction based approach to the Kochen-Specker theorem', submitted to *Journal of Physics A*, quant-ph/9912018.

Hamilton, J, Isham, C. and Butterfield, J.: 2000, 'A topos perspective on the Kochen-Specker theorem: III. Von Neumann Algebras as the Base Category', forthcoming *International Journal of Theoretical Physics*, March 2000, quant-ph/9911020.

Isham, C. and Butterfield, J.: 1998, 'A topos perspective on the Kochen-Specker theorem: I. Quantum states as generalised valuations', *International Journal of Theoretical Physics*, **37**, pp. 2669–2733. quant-ph/980355.

Isham, C. and Butterfield, J.: 2000, 'Some Possible Roles for Topos Theory in Quantum Theory and Quantum Gravity', forthcoming in *Foundations of Physics*, gr-qc/9910005.

Lawvere, W. and Shanuel, S.: 1997, *Conceptual Mathematics: a First Introduction to Categories*, Cambridge University Press, Cambridge.

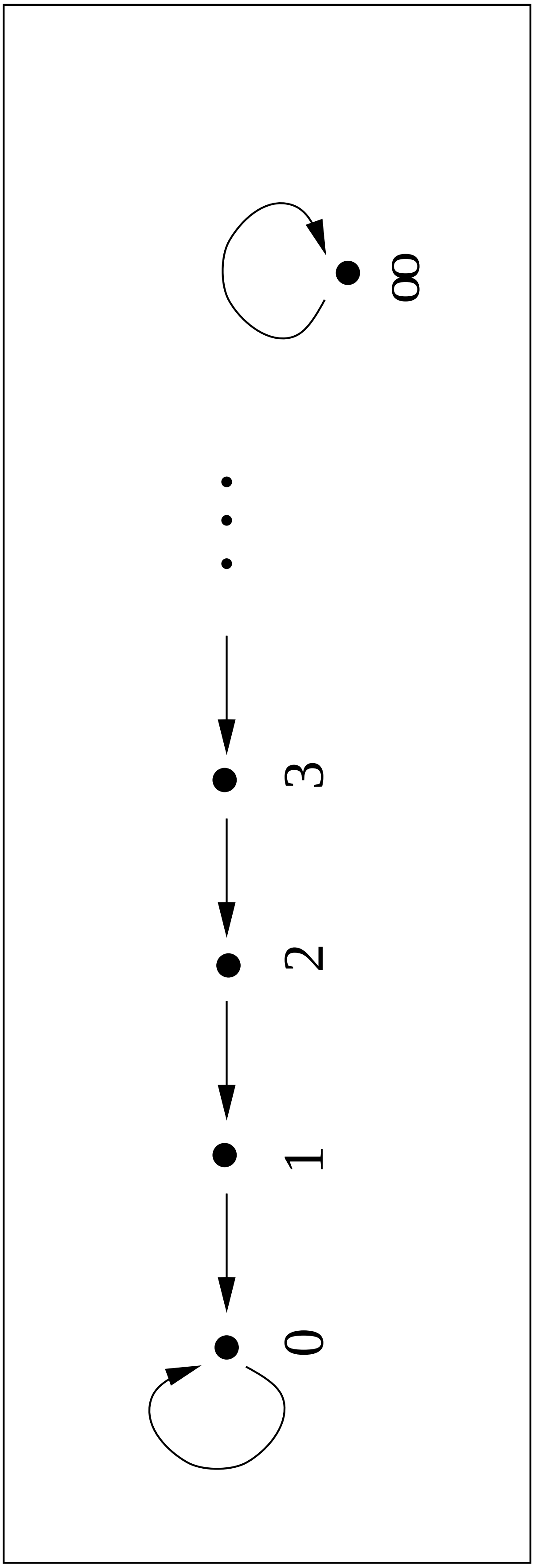

Figure 18: Ω for $\mathcal{E}$

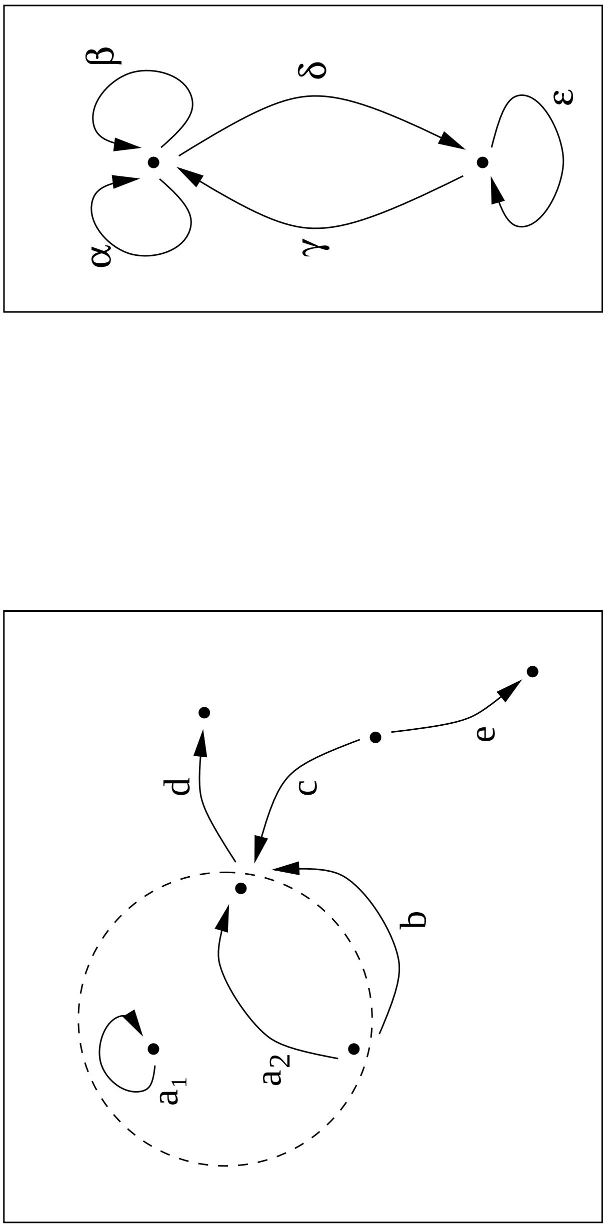

Figure 2: A graph, a subgraph and Ω for $\mathcal{G}$

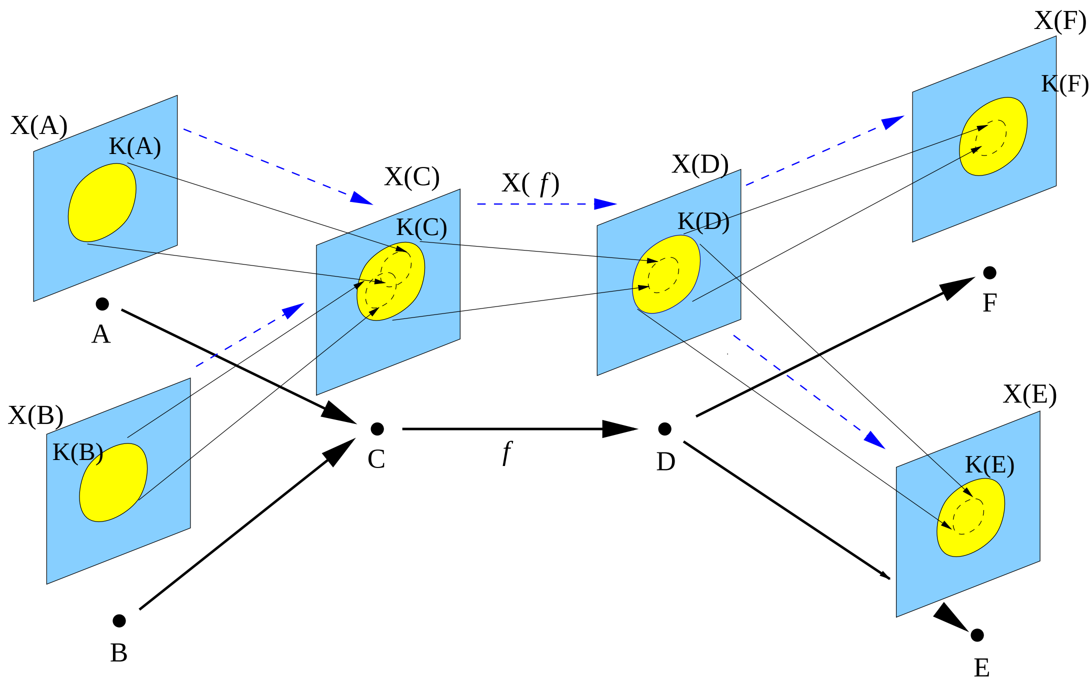

Figure 3: A presheaf **X**, and a subobject **K** of **X**