[241]

Theory of Concepts*

Erich Rast

Key Notions and Problems

Concept theories draw on a rich tradition, ranging from Plato and Aristotle over Leibniz to Frege. Two key aspects of a theory of concepts need to be distinguished. (i) The cognitive aspect regards the role of concepts in cognition and how these enable an epistemic agent to classify and categorize reality. A concept system is sometimes considered the cornerstone and starting point of a 'logic of thinking.' (ii) From a metaphysical point of view, concept theory must provide an explanation of the ontological status of universals, how these combine, whether there are different modes of predication, and what it means in general for an object to fall under a concept. Both aspects will be addressed in what follows. The survey starts with a brief overview of selected problems and positions.

The Demarcation Problem. There is no general agreement in the literature on what a concept is. Sometimes 'concept' is more or less used as a synonym for 'property', but many authors use it in a more specific sense, for example as standing [242] for unsaturated entities whose extensions are sets and classes (Frege), for Fregean senses (Church), or for abstract objects (Zalta). One goal shared by many authors, despite terminological differences, is to carve out the differences between closely related notions such as concepts, properties, abstract objects, Leibnizian concepts, or Fregean senses and make these notions more precise.

Nominalism, Realism, Cognitivism. A particular object is said to fall under a singular or individual concept and likewise a group of objects sharing some common trait is said to fall under a general concept. Being sorts of universals, different stances towards general concepts may be taken: According to strict nominalism there are only particulars; quantification over predicate expressions is not allowed at all or very limitedly. In this view general concepts do not exist in reality although they might play a role as thinking

*Published in: Hansson, Sven Ove & Hendricks, Vincent (eds.): *Introduction to Formal Philosophy*, Cham, Switzerland: Springer 2019, pp. 241-250. This is an earlier draft version with the original pagination in square brackets. Please quote the original article.

devices. In contrast to this, according to realism predicates denote universals either directly or whenever the predicate has been nominalized. There *are* universals in the sense that one may fully quantify over them although they might not be considered to exist in the narrow sense. Cognitivism is a mixed position. In this view, there are universals but only insofar as they are represented (or representable) by mental states.

Intensionality, Hyperintensionality, Contradictory Concepts. Having a heart and having a liver are often given as an example of two different concepts with the same extension. Modal logics have been used to account for this difference. Normal possible worlds semantics does not, however, provide the means to distinguish two different mathematical concepts with the same extension from each other. For example, two different ways of describing an equiangular triangle will determine the same set of objects in all possible worlds. To tackle this problem a stronger form of intensionality known as hyperintensionality is needed. Moreover, a person might erroneously believe that 37 is a prime number while not believing that $21 + 16$ is prime, might erroneously believe that $\sqrt{2}$ is a rational number, or might muse about round squares. To represent irrational attitudes and impossible objects a logic must in one way or another allow contradictory statements. Since in classical logic any formula can be derived from a contradiction (ex falso quod libet) a paraconsistent logic is needed; such a logic allows one to derive some, but not arbitrary consequences from a contradiction.

Similarity. A concept may be more or less similar to other concepts. For example, the concept of being a chair is similar to the concept of being a stool and both of them are more similar to each other than any of them is to the concept of being the back of a horse. From a cognitive perspective it is desirable to have a concept theory that allows for a measure of similarity between concepts and the objects falling under them.

Typicality. Typically chairs have four legs, but some have less. Typically birds can fly, but penguins cannot fly. How can this typicality be accounted for?

[243]

Preliminaries of Logical Concept Theory

In order to formulate a broadly-conceived logical theory of concepts it is necessary to quantify over concepts or corresponding abstract objects. Unless a very strict nominalism based on first-order logic is defended this naturally involves the use of second-order logic. For this reason results from mathematical logic need to be taken into account when developing a logical theory of concepts, some of which are addressed in what follows.

Henkin Models and Standard Models. There are two kinds of models for higher-order logic. In a standard model, first-order variables range over a domain D, second-order variables over $\mathcal{P}(D)$ for predicates and $\mathcal{P}(D_1 \times \cdots \times D_n)$ for n-ary relations, third-order predicate variables over

Stratification: Formula ϕ is homogeneously stratified iff there is a function $f(.)$ that maps terms and formulas of the language to natural numbers such that for any atomic formula $P(x_1, \ldots, x_n)$ in ϕ, $f(P) = \max[f(x_i)] + 1$ and $f(x_i) = f(x_j)$ for $1 \leq i, j \leq n$.

$$\exists F \forall \vec{x}[F(\vec{x}) \leftrightarrow \phi(\vec{x})] \quad \text{(Scheme A)}$$
$$\exists F \forall \vec{x}[F(\vec{x}) \leftrightarrow (G(\vec{x}) \wedge \phi(\vec{x}))] \quad \text{(Scheme B)}$$

Conditions: (I) $\vec{x} := x_1, \ldots, x_n$ are free in ϕ, i.e. bound in the whole scheme; (II) F is not free in ϕ, i.e. not bound in the whole scheme; (III) ϕ is homogeneously stratified.

① Unrestricted Comprehension: Scheme A + I

② Predicative Comprehension: Scheme A + I, II, III

③ Separation Axiom: Scheme B + I, II

Box 1: Comprehension Schemes and Stratification

$\mathcal{P}(\mathcal{P}(D))$, and so on. In a Henkin model (general model), only a fixed subset of the powerset is chosen respectively. So for instance the quantifier in $\forall F[F(a)]$ ranges over a fixed subset of $\mathcal{P}(D)$. Higher-order logic with Henkin models is essentially a variant of many-sorted first-order predicate logic (Henkin, 1950). It is complete, compact and the Löwenheim-Skolem theorems hold in it, but does not allow one to define certain mathematical structures categorically, i.e. in a way that is unique apart from differences captured by the notion of an isomorphism between models. In contrast to this, higher-order logic with standard models is not complete, not compact, and the Löwenheim-Skolem theorems do not hold in it. Lack of a full-fledged proof theory is compensated by the ability to categorically define important concepts such as countable vs. uncountable domains, quantifiers like 'most', and well-foundedness conditions. The distinction between higher-order logic and second-order logic with standard models is less important, since the former can be reduced to the latter without significant loss of expressivity (Hintikka, 1955). For this reason many authors focus on second-order logic.

[244] *Logical Paradoxes and Comprehension.* Given some condition expressible in a formal language, what concepts are there? One way to answer this question is by specifying a comprehension scheme. Unrestricted comprehension asserts that there is a concept corresponding to any condition ϕ that can be formulated in the language (see Box 1, Principle ①). It allows one to introduce Russell's paradox of predication, the analogue to the well-known set-theoretic paradox. Take the predicate $P(x)$ that is not predicable of itself and is defined as $\neg x(x)$. Choosing $\phi := \neg x(x)$

and existential instantiation allows one to derive $\forall x[P(x) \leftrightarrow \neg x(x)]$ and by universal instantiation the contradiction $P(P) \leftrightarrow \neg P(P)$. Different provisions to avoid such inconsistencies lead to higher-order logics with varying expressive power that reflect different stances towards nominalism, cognitivism, and realism.

Predicativity vs. Impredicativity. A definition is impredicative iff it quantifies over a collection of objects to which the defined object belongs; otherwise it is predicative. Some mathematicians like Poincaré, Weyl, and Russell himself held the view that paradoxes arise because a logic with unrestricted comprehension allows for impredicative definitions. As a solution, the logic is made predicative. One way to achieve this is by assigning an order to all variables and prescribe that in any atomic formula $P(x_1, \ldots, x_n)$ in a condition ϕ formulated in the language the order of all x must be lower than the order of P (see Box 1, Principle ② and *stratification*). This makes $\neg x(x)$ ungrammatical. Church's influential Simple Type Theory (STT) is another way to define a predicative higher-order logic. Every term has a type with corresponding domain. Starting with finitely many base types, infinitely many compound types can be built. If α and β are types, then $(\alpha\beta)$ is the type of a function that takes an object of type β and yields an object of type α, where β and α may themselves be compound types. Predicates and relations are represented by several functions. This is called Currying or Schönfinkelization. For example, a unary predicate P is of type $(\sigma\iota)$, indicating a function that takes a term of type ι and yielding a truth-value of type σ, a second-order predicate is of type $(\sigma(\sigma\iota))$, and so on. (Another notation which was popularized by Montague uses e for objects, t for truth-values, and the order is reversed.)

Impredicativity does not automatically lead to paradoxes. On the contrary, many useful mathematical concepts such as the induction principle used for defining natural numbers are impredicative. For this reason some conceptual realists opt for impredicative second-order logics that give rise to larger mathematical universes. In these logics comprehension is restricted less radically than in predicative ones (see e.g. Box 1, Principle ③) or full comprehension is combined with a limited substitution principle in order to gain more expressivity while avoiding the paradoxes. The downside is that it is harder to ensure consistency in such systems than in purely predicative logics.

Philosophical Relevance. First-order logic and predicative higher-order logic with Henkin models reflect a strict nominalist stance as has been defended by Lésniewski, for example. Predicative higher-order logic with standard models may also be considered nominalist in spirit, because predicative comprehension reduces the existence of general concepts to conditions explicitly given in the language. [245] In contrast to this, impredicative higher-order logics with standard models clearly reflect a realist stance. More fine-grained distinctions can be found in Cocchiarella (1994, 2007).

Concepts as Abstract Objects

Possibilism. While the conceptual realist wants to talk about concepts it would be implausible to claim that concepts exist in the same sense as ordinary objects. Therefore, many conceptual realists distinguish, pace Quine, between quantification as a means of counting and quantification as a means of asserting existence. A logic in which non-trivial properties can be ascribed to nonexistent objects is possibilist or Meinongian, where the latter term is often used for metaphysical theories that allow one to talk about contradictory objects. In a classical setting, possibilism can be obtained by introducing two sorts of quantifiers. Actualist quantifiers are mere means of counting and run over the total domain, whereas possibilist quantifiers additionally assert existence and run only over a subset of the total domain. Alternatively, a unary existence predicate $E(x)$ may be introduced to which possibilist quantifiers are relativized, for instance $\forall^* xA := \forall x[E(x) \rightarrow A]$ and $\exists^* xA := \exists x[E(x) \wedge A]$.

Nominalization. One positive answer to the problem of universals is to assert that we cannot only quantify over concepts but are also able to talk about concepts like *being nice* as objects. Sometimes λ-abstraction is thought to fulfill this purpose. Semantically, a term of the form $\lambda x.P(x)$ is interpreted as the function that with respect to an assignment g takes an a within the domain of x and whose result is the same as $P(x)$ evaluated with respect to the modified assignment g' that is the same as g except that $g'(x) = a$. One might then consider $\lambda x.P(x)$ to stand for *being nice* if P stands for the predicate *nice*. However, λ-terms can be used instead of relations (as in STT) and the converse transformation is also possible in a logic with both functions and relations, and so λ-abstraction might not be considered a tool for nominalization understood in the narrow sense. Abstract object theory (Zalta, 1983) and alternative ontologies such as trope theories (Rapaport, 1978; Castañeda, 1989; Mormann, 1995) provide more elaborate nominalization mechanisms. Differing considerably in details and terminology, generally in these approaches nuclear and extranuclear properties are distinguished from each other (Parsons, 1980), where the former are being constitutive of an object and the latter are not, and two different modes of predication are available: An object, which does not have to be concrete or existent, encodes a property if the property takes part of a description or listing of the object's essential features whereas it exemplifies a property if it has the property accidentally. For example, in bundle trope theories an object encodes a property if its constituting bundle of properties (viz., property moments also sometimes called qualitons) contains the property and exemplifies a property if it stands in a designated relation to the property. A concept is, in this view, a nonexistent non-concrete bundle of primitive properties or property moments. Analogously, in abstract object theory aF stands [246] for the fact that the abstract (nonexistent) object a encodes

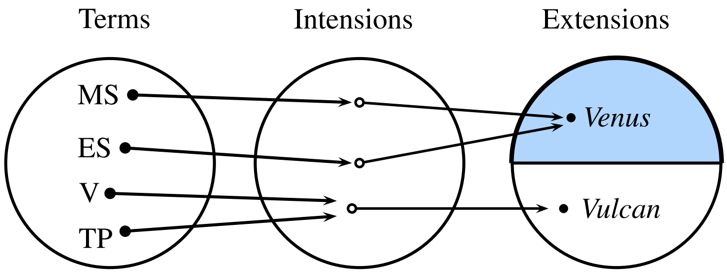

Figure 1: Intensional versus extensional identity in a possibilist intensional logic. ES='evening star', MS='morning star', V='Vulcan', TP='the planet between Mercury and Sun.'

property F. Care must be taken to restrict the range of properties that can be encoded. For example, forming an abstract object *existent red sphere* must either be disallowed or the existence-entailing predicate 'existent' must be interpreted in a derived, non-literal way in this construction.

Concepts and Intensionality

Modal Concepts and Intensionality. Modal operators may be added to higher-order logic in the same way as they are added to first-order logic, which in the second-order setting allows one to precisely express philosophical positions about the modal properties of concepts. For example, Anti-Essentialism may be expressed by adding the following axiom:

$$\forall F[\exists x \Box F(x) \rightarrow \forall x \Box F(x)] \tag{1}$$

which may be paraphrased as "if an object has an essential property, then any object has this essential property."

Hyperintensionality. Inspired by Frege's informal distinction between the sense and the denotation of an expression, there is a tradition of hyperintensional logics in which the following Axiom of Extensionality does not hold:

$$\forall F \forall G(\forall \vec{x}[F(\vec{x}) \leftrightarrow G(\vec{x})] \rightarrow \forall H[H(F) \rightarrow H(G)]) \tag{2}$$

This axiom states that if exactly the same objects fall under two concepts, then the concepts are identical in Leibniz' sense of having the same properties. Despite considerable differences in detail, hyperintensional logics generally invalidate this axiom by interpreting expressions over a domain of fine-grained intensions, which are in turn mapped to their extensions by an extension function (Muskens, 2007). Consequently, two notions of identity are available in such a logic: coarse-grained extensional identity and fine-grained intensional identity interpreted over intensions (Fig. 1). By interpreting functions and operators standing for notions like de dicto belief over intensions it is possible to distinguish having a heart from having a liver and deal with ordinary cases of referential opacity like Frege's Morning–Evening Star example. Additionally, strong intensions allow one to represent attitudes that are not closed under logical consequence, i.e. someone's believing that 37 is prime while not believing that 21+16 is prime.

[247] *Contradictory Concepts.* Representing irrational attitudes or contradictory concepts like being a round square requires substantial changes to the underlying logic. In a modal logical setting sometimes impossible worlds are introduced. At an impossible world 'anything goes'; arbitrary formulas, including contradictions, may be true at such a world by mere syntactic assignment. Another approach based on seminal work by Asenjo, da Costa, Anderson and Belnap is to use a 3-valued logic such as LP or RM3. These logics are paraconsistent and allow a contradictory formula to have a designated truth value that is interpreted as 'both true and false.' Paraconsistent logics have also been proposed as a way of dealing with the paradoxes, allowing the logic to mirror the philosophical position that there *are* real paradoxes and our talk about them is meaningful (Dialetheism).

The logical aspects of concept theory mentioned so far are well-known, but are not commonly combined into one all-encompassing metaphysical theory. Most authors focus on some of these aspects, such as how they can be used to answer the problem of universals, or logical reconstructions of historical positions such as Leibniz' Concept Calculus or Platonic Forms. References to further work are given in Section 6.

Geometrical Approaches

In this section some promising alternatives to the logical approach shall be mentioned, which are not metaphysical in the narrow sense. These broadly-conceived geometrical concept approaches fare particularly well with issues related to the cognitive aspects of concepts such as vagueness, typicality, and similarity and can either be combined with, or are thought to complement, logical theories.

Typicality. In a qualitative approach a preorder relation (preference relation) between all objects falling under a concept can be used to order objects falling under a given concept according to their typicality. The center represents a prototype and the nearer an object is to the center the more typical it is (Fig. 3a). In a logical setting this kind of typicality can be expressed in Preference Logics and related descendants of Lewis' Conditional Logic. Quantitative accounts induce a similar ordering by assigning a degree of typicality as a real number between 0 and 1 to each object as dependent on the concept it falls under, and despite some differences the two approaches can for many practical purposes be translated into each other. There are interconnections of these basic forms of typicality to non-monotonic logics for default reasoning, belief revision, ϵ-entailment, and plausibility and possibility measures.

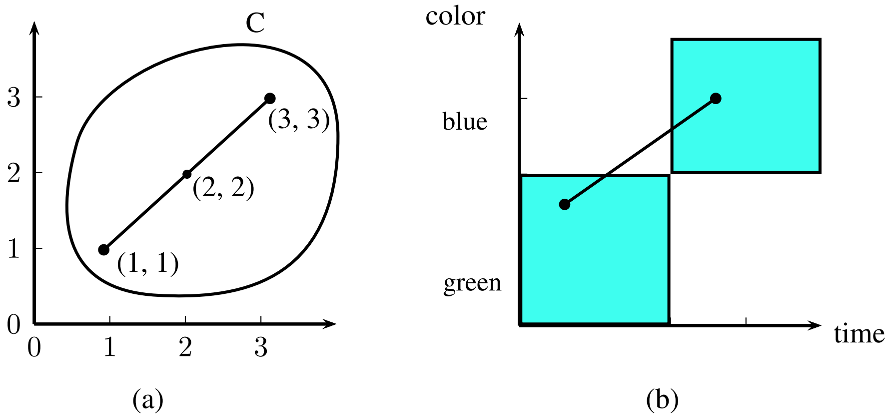

Figure 2: (a) Convexity: for example for $t = \frac{1}{2}$ the point $\frac{1}{2} \cdot \langle 1, 1\rangle + (1 - \frac{1}{2}) \cdot \langle 3, 3\rangle = \langle \frac{1}{2}, \frac{1}{2}\rangle + \langle \frac{3}{2}, \frac{3}{2}\rangle = \langle 2, 2\rangle$ is also in C. (b) Goodman's concept *grue* is not convex.

Conceptual Spaces. Gärdenfors (2004) proposes to model concepts not just on a symbolic, but also on a geometrical level. A conceptual space is an n-dimensional metric space with n quality dimensions, each of which represents a basic quality like *height*, *width*, *hue*, *saturation*, or *loudness*. A distance function allows for measuring the distance between any two points in such a space. In the simplest case of familiar n-dimensional Euclidean space this distance measure between two points $x = \langle x_1, \ldots, x_n\rangle$ and $y = \langle y_1, \ldots, y_n\rangle$ is defined as

$$d_E(x, y) = \sqrt{\sum_{i=1}^{n} w_i (x_i - y_i)^2} \tag{3}$$

[248] where w_i represents the weight of the respective quality dimension. More general topological definitions of spaces allow for an adequate treatment of purely qualitative dimensions. Generally speaking, in a conceptual space objects are represented as vectors $x = \langle x_1, \ldots, x_n\rangle$ and concepts by regions in the space. Similarity between two objects in a conceptual space is defined as a function of their distance.

Gärdenfors has conjectured that natural concepts should be represented by convex regions. A region C of a space S is convex iff for any two points $x, y \in C$ any point $tx + (1 - t)y$, where $0 \leq t \leq 1$, on the line segment $\overline{xy}$ between x and y is also in C (Fig. 2a). One advantage of this assumption is that every convex region has a center, which may be interpreted as a prototypical object falling under the concept. Taking these centers $p_1, \ldots, p_k$ as starting points, concepts C_i can be defined around them by partitioning the space such that for each point $x \in C_i$, $d(p_i, x) \leq d(p_j, x)$ if $i \neq j$. The result is called a Voronoi diagram (Fig. 3b). The closer a point is to the center p_i

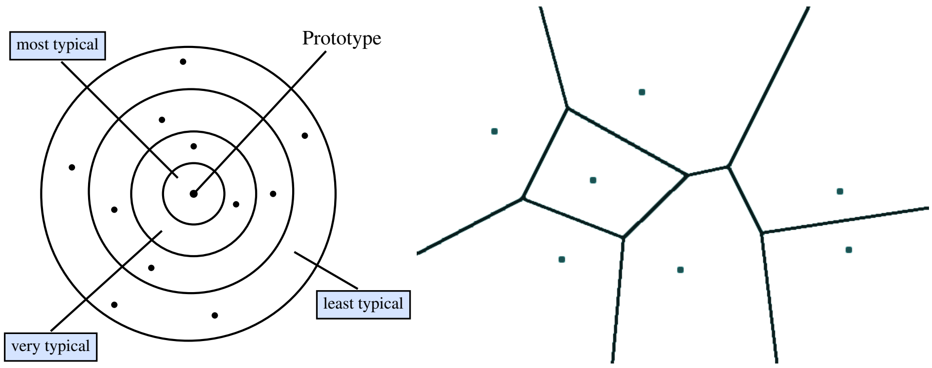

Figure 3: (a) Typicality expressed as a preference ordering. (b) A Voronoi diagram with seven centers.

of its concept C_i in such a partitioning, the higher is the degree of typicality of the object it represents. The convexity condition has also been taken as a first step toward distinguishing between natural and non-natural concepts. For example, with 'standard' quality dimensions Goodman's artificial concept *grue*, which is true of green objects before some point in time and of blue ones afterwards, is represented by a non-convex region (Fig. 2b). However, this solution depends on criteria for finding natural quality dimensions, as a natural concept may be turned into a non-natural one by changing the underlying dimensions and vice versa.

Formal Concept Analysis. In formal concept analysis a set M of attributes is associated with a set of objects G by a binary relation $I(x, y)$ read as "object x [249] has attribute G", where the triple $\langle M, G, I\rangle$ is called a context. Such a context may be thought of as a table with objects as rows and attributes as columns and a mark at the row-column intersection if the object at that row has the respective attribute. A formal concept is then a pair $\langle A, B\rangle$ of subsets $A \subseteq G$ and $B \subseteq M$ such that all objects in A share all the attributes in B. The formal concepts of a context can be ordered by a relation $(A, B) \leq (C, D)$ which is true iff $A \subseteq C$, false otherwise. Ordering all concepts in a context yields a *lattice* structure in which the least specific concept is at the bottom and the most specific one is at the top. Various methods and algorithms based on this representation have been used for data mining, machine learning, discovering new relationships between concepts, concept visualization, explaining human concept acquisition, and models of concept change.

Further Reading

Andrews (2002) contains an introduction to type theory; reprints of original articles can be found in Benzmüller et al. (2008). Shapiro (1991) is a comprehensive treatment of second-order logic. Burgess (2005) discusses predicative and impredicative foundations of arithmetics with a focus on Frege. Metaphysical implications of different comprehension schemes are discussed at length in Cocchiarella (1994, 2007). Priest (2005) is a modern defense of possibilism and dialetheism; it may serve as a reference for further literature. Zalta (1983) is the main work on abstract object theory and contains reconstructions of Platonic Forms and Leibniz' Concept Theory; many refinements can be found in Zalta's more recent works. Gärdenfors (2004) is the seminal work on Conceptual Spaces. Ganter and Wille (1998) and Ganter et al. (2005) lay out formal concept analysis in a rigid manner. [250]

References and Recommended Readings

Asterisks (*) indicate recommended readings.

* Andrews, P. B. (2002). *An Introduction to Mathematical Logic and Type Theory: To Truth through Proof*. Academic Press, New York. [Contains an introduction to higher-order logic suitable for beginners].

Benzmüller, C., Brown, C. E., Siekmann, J., and Stratman, R., editors (2008). *Reasoning in Simple Type Theory*, number 17 in Studies in Logic, London. College Publications.

Burgess, J. P. (2005). *Fixing Frege*. Princeton University Press, Princeton, N J / Oxfordshire.

Castañeda, H.-N. (1989). *Thinking, Language, Experience*. University of Minnesota Press, Minneapolis, Minn.

Cocchiarella, N. B. (1994). Philosophical perspectives on formal theories of predication. In Guenthner, F. and Gabbay, D., editors, *Handbook of Philosophical Logic*, volume 4, pages 253–326. Kluwer.

* Cocchiarella, N. B. (2007). *Formal Ontology and Conceptual Realism*. Synthese Library Vol. 339. Springer, Dordrecht. [Intermediate-level text that lays out fine-grained philosophical distinctions between systems of higher-order logic].

Ganter, B., Stumme, G., and Wille, R., editors (2005). *Formal Concept Analysis*. Springer.

Ganter, B. and Wille, R. (1998). *Formal Concept Analysis: Mathematical Foundations*. Springer, Berlin / Heiderlberg / New York.

* Gärdenfors, P. (2004). *Conceptual Spaces: The Geometry of Thought*. MIT Press, Cambridge, MA. [Reference work on the theory of conceptual spaces].

Henkin, L. (1950). Completeness in the theory of types. *The Journal of Symbolic Logic*, 15(2):81–91.

Hintikka, J. (1955). Reductions in the theory of types. In *Two Papers on Symbolic Logic*, volume 8 (1955), pages 57–115. Acta Philosophica Fennica.

Mormann, T. (1995). Trope sheaves: A topological ontology of tropes. *Logic and Logical Philosophy*, 3:129–150.

Muskens, R. (2007). Intensional models for the theory of types. *Journal of Symbolic Logic*, 72(1):98–118.

Parsons, T. (1980). *Existence.* Yale University Press, New Haven.

* Priest, G. (2005). *Towards Non-Being: The Logic and Metaphysics of Intentionality.* Clarendon, Oxford. [Seminal recent work on non-existence].

Rapaport, W. J. (1978). Meinongian theories and a russellian paradox. *Noûs*, 12(2):153–180.

Shapiro, S. (1991). *Foundations without Foundationalism: A Case for Second-Order Logic.* Clarendon Press, Oxford.

* Zalta, E. (1983). *Abstract Objects.* D. Reidel, Dordrecht, Boston, Lancaster. [Older, yet important work on abstract object theory].