

Revising Concepts

Renata Wassermann

Institute for Logic, Language and Computation*
University of Amsterdam
email: renata@wins.uva.nl

Abstract

In this paper we present a model for revising concepts using Grove's model for belief revision [Gro88]. We model concepts and objects by means of conceptual spaces [Gär98] and show how to deal with prototype effects. This extends the model presented in [Leh98].

1 Introduction

Belief revision (for an overview, see [Gär88, GR95, Han98]) deals with the problem of how to accommodate new assertions into an existent body of knowledge. Traditionally, the body of knowledge is represented by a belief set, a set of formulas closed under logical implication. A belief set can also be represented by the set of possible worlds where all its formulas hold.

The idea of this paper is to apply methods developed in the context of belief revision in order to accommodate new information about concepts. If we think of the representation of a belief set in terms of possible worlds, it contains the set of worlds that the agent holds for possible, given his (partial) knowledge. We will be talking of partial descriptions of an object, instead of descriptions of the world, and of sets of possible objects instead of possible worlds. The intuition behind it is that, given a description, there is a set of objects that the agent holds for possible, that is, a set of objects that could be the one being described.

For the sake of simplicity, we will concentrate on the semantical aspects of the definitions. Concepts are defined by the properties an object must have in order to be classified as an instance of the concept.

In section 2 we briefly present Grove's model for belief revision [Gro88]. In section 3 we show how to change perspective and model concept revision using Grove's ideas. In section 4 we show how the model can be enriched to deal with prototype effects and in section 5 we present some properties of the nonmonotonic inference notion obtained. Finally, in section 6 we conclude and point towards future work.

2 Sphere Models

A belief set can be represented by the set of worlds consistent with the beliefs in it, that is, by the set of worlds that the agent holds for possible given his beliefs. Grove [Gro88] has given a model for belief revision based on a family of spheres around a given belief set K . Each sphere is a possible weakening of K , that is, the set of possible worlds obtained by giving up some information in K . Each sphere is also called a fallback.

Formally, let \mathcal{L} be a language, M be the set of all possible worlds (valuations of the language), Cn the classical consequence operation. A set $K \subseteq \mathcal{L}$ is a belief set if and only if $K = Cn(K)$. Given a belief set K , $[K]$ is the set of possible worlds compatible with K . Given a subset A of M , $t(A)$ is the set of formulas that hold in all worlds of A . Grove defines a family of spheres centered on $[K]$ as a collection S of subsets of M such that:

*Plantage Muidersgracht 24, 1018-TV, Amsterdam, The Netherlands

- (S1) If $U, V \in S$, then $U \subseteq V$ or $V \subseteq U$.
- (S2) $[K]$ is the \subseteq -minimum of S .
- (S3) M is the \subseteq -maximum of S .
- (S4) If α is a sentence in \mathcal{L} and there is any sphere in S intersecting $[\alpha]$, then there is an inclusion minimal sphere intersecting $[\alpha]$.

Grove defines the revision of a belief set K by a sentence α given a family of spheres S centered on $[K]$ as the intersection $f_S(\alpha)$ of $[\alpha]$ and the smallest sphere of S such that the intersection is not empty. He shows that if one defines a revision operation as $K *_G \alpha = t(f_S(\alpha))$, then $*_G$ satisfies the 8 AGM postulates for belief revision.

3 A Change of Perspective

In this section we will be concerned with revising not assertions about the world, but a description of a particular object. Such a description may be given by some formula of a description language. We will not go into details about the language here, we will only assume that it contains the connective \wedge and that this connective is interpreted as: if $[\alpha]$ and $[\beta]$ are the sets of objects satisfying the descriptions α and β , then $[\alpha \wedge \beta] = [\alpha] \cap [\beta]$.

Since we are interested in revision of concepts, we would like to change perspective. We are dealing with (partial) representations of objects. Instead of associating to a concept description a set of worlds where the concept satisfies that description (like in [Leh98]), we find it more intuitive to associate with the description a set of **objects** satisfying it. These are all the objects the agent takes for possible given the description.

Following Gärdenfors [Gär98], we will consider multidimensional spaces for representing concepts. Each dimension is a quality, roughly equivalent to an attribute name in a frame structure. An object is a point in a multidimensional space, representing the values it has for each attribute. A concept is a convex region of a multidimensional space.

By a concept, we mean what someone believes to be its definition. Suppose I learnt at school (and believed it!) that water is a substance with no color, no smell, no taste, with formula H_2O and that only occurs on the Earth. Then my concept of water includes all those objects that have no color, no smell, no taste, with formula H_2O and that only occur on the Earth.

Each concept comes equipped with a family of fallbacks, that represent possible “weakenings” of the concept, that is, sets of objects that satisfy only part of the description. As in the case of Grove’s models for belief revision, there is an assumption here that the fallbacks can be completely ordered. This is a simplification. The general case, where the fallbacks are not necessarily nested, will be addressed in future work.

Suppose now that I read (and believe) that water was found on the Moon. I want to revise my concept of water to accommodate the new belief, but this should not change my belief that water has formula H_2O . My new concept of water should be as close to the original as possible. I should choose from the objects existent on the Moon as well as on the Earth the ones that are the most similar to the ones I held for possible before.

In figure 1 we see the set of objects that satisfy the new concept of water. The set is the intersection of the set of objects present on the Moon and on the Earth and the smallest compatible fallback of the old concept.

4 Prototype Effects

One of the main characteristics of human reasoning is the ability of jumping to conclusions even on the absence of complete information. People can communicate without filling in all the details about the object they are talking about. When I hear that someone has a bird, I imagine a more or less typical bird, imagine it flies, etc. But of course I may be wrong, the person may be talking about a bird that does not fly. This tendency to pick up most typical objects has been studied in prototype theory ([Ros73, Lak87]). Gärdenfors has also addressed prototype effects on conceptual spaces.

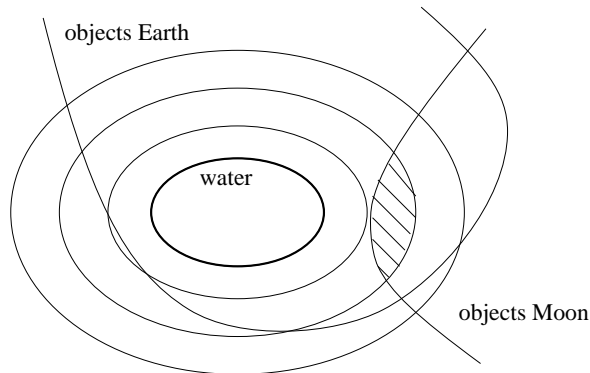


Figure 1: Revising the concept of water

The use of multidimensional spaces instead of frame-like structures makes some aspects of concept descriptions clearer. One can talk about concepts being close to each other, and most important, about objects being more or less central in a concept. Typical objects are more central than others. Desk-chairs are more typical than rocking chairs and should be represented by points near the center of the region representing the concept chair.

Gärdenfors defines the prototypical region of a concept as a sphere around the center of the concept. The radius of the sphere has to be determined empirically. One can talk about nonmonotonic properties of a concept as those that are true of every object in the prototypical region. If the only information one has about an object is that it is a bird, it is plausible that it flies, since typical birds fly. If one then learns that the bird is a penguin, the prototypical region will move to a less central position containing birds that do not fly.

In [Leh98], Lehmann introduces a formal model for prototypical reasoning (called stereotypical reasoning there). It is assumed that there is a collection of stereotypes, that are sets of possible worlds. For each belief set K (also seen as a set of worlds), one has to look for the stereotype S^K that covers it best. One can use then the intersection of both sets for drawing conclusions. Lehmann defines some ways of finding the best stereotype using distance functions and shows some interesting properties of the inference function derived from it. But the paper only deals with the case where there is a stereotype such that its intersection with the belief set is non-empty.

If we think of the modeling in terms of sets of possible objects, it is not difficult to see that Lehmann's operation can be seen as a counterpart of the expansion of a belief set. The information coming from the stereotype is simply added to the description one had and the new inference relation is given by: $C(K) = Cn(K \cup S^K)$. Following the analogy, the case where the intersection of $[K]$ and $[S^K]$ is empty should be the counterpart of a revision operation.

We will assume that not only the sets representing concepts have a family of fallbacks, but also the set of prototypical elements. The revision of a concept is then a revision of the concept itself and the revision of the prototypical region.

The family of fallbacks around a prototype has the prototypical region as its minimum and the concept as its maximum.

This can be more easily seen by means of an example. Suppose we have the concept tiger defined by means of properties like being a vertebrate, having a certain DNA structure, etc. Prototypical tigers are big, yellow with black stripes, carnivore, have four legs, etc. We want to revise this concept due to new information acquired. Five situations are possible, as can be seen in figure 2.

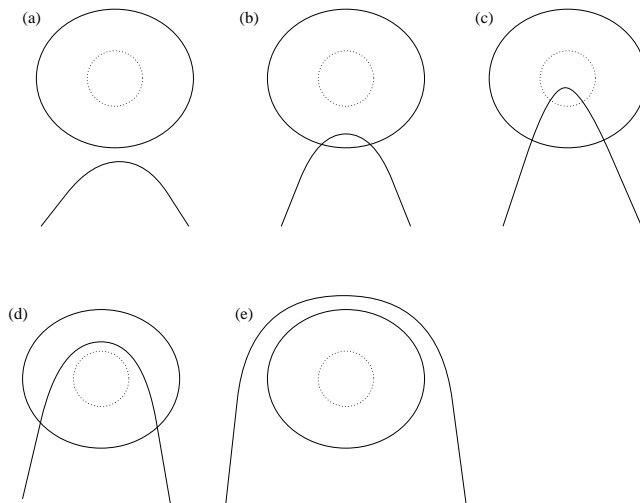


Figure 2: Revising prototypes

In situation (a), the new information is completely incompatible with even the essential properties of tigers. Suppose the new input is that the object is invertebrate. There is no such thing as an invertebrate tiger. In this case, the new concept assumed should be the intersection of the invertebrate objects with the inclusion minimal fallback of tiger that is compatible to it. This is equivalent to a revision of the concept of tiger by the property of being invertebrate. The resulting concept is that of invertebrates that look as much as a tiger as possible. There is no information about the new prototypical region.

In situation (b), the new information is compatible with tiger but not with prototypical tiger. Suppose it is that the object has three legs. The new concept is that of three-legged tiger, that is a subset of tiger. The new prototypical region is the intersection of the three-legged objects with the minimal compatible fallback of prototypical tiger, that is, those three-legged tigers that are as close to prototypical tigers as possible (in that they are big, yellow with black stripes, etc.).

In situation (c), the new information is compatible with tiger as well as with prototypical tiger. It could be that the object is female. The concept of female tiger is the intersection of tiger and female and the prototypical female tiger is the intersection of prototypical tiger with female, that is, those female tigers that are big, yellow with black stripes, etc.

In situation (d), the new information is valid for all prototypical tigers, but not for all tigers. It could be that the object has four legs. In this case, the new concept is that of a four-legged tiger, that is the intersection of tiger and four-legged. The prototypical region stays the same, prototypical four-legged tigers are prototypical tigers.

Finally, in situation (e), the new information is already valid for all tigers, for example, that they are vertebrates. Nothing changes, vertebrate tigers are tigers.

5 Formal Properties

In this section we describe some properties of the inference relation derived from prototypical reasoning.

We formalize the idea of prototypical region by means of a selection function. Intuitively, the function p chooses from a set of objects those which are the most typical.

Definition 5.1 *Let p be a function from sets of objects to sets of objects such that for all α , $p([\alpha]) \subseteq [\alpha]$. The elements of $p([\alpha])$ are called prototypical α -objects.*

Following Gärdenfors [Gär98] and Lehmann [Leh98] we say that β is a nonmonotonic property of α if and only if all prototypical α -objects have property β , that is:

Definition 5.2 $\alpha \sim \beta$ iff $p([\alpha]) \subseteq [\beta]$, where $p([\alpha])$ is the set of prototypical α -objects.

[Gab85] proposed that any inference relation should satisfy at least three properties:

1. $\alpha \sim \alpha$ (reflexivity)
2. if $\alpha \wedge \beta \sim \gamma$ and $\alpha \sim \beta$, then $\alpha \sim \gamma$ (cut)
3. if $\alpha \sim \beta$ and $\alpha \sim \gamma$, then $\alpha \wedge \beta \sim \gamma$ (cautious monotonicity)

In [KLM90], an inference relation is called cumulative if and only if it satisfies the three properties above and:

4. if $[\alpha] = [\beta]$ and $\alpha \sim \gamma$, then $\beta \sim \gamma$ (left logical equivalence)
5. if $[\alpha] \subseteq [\beta]$ and $\gamma \sim \alpha$, then $\gamma \sim \beta$ (right weakening)

Lehmann defines a way of choosing the prototype of a concept using a distance function and shows that the inference function obtained is cumulative [Leh98]. The same result can be obtained with a very weak assumption on the function p :

(*) For all sets of objects X and Y , if $p(X) \subseteq Y \subseteq X$, then $p(X) = p(Y)$.¹

Theorem 5.3 Let p be as in 5.1. If p satisfies (*) then the inference function defined as in 5.2 is cumulative.

To see what it means to have p satisfy (*), consider the example illustrated by figure 2(d). The information that the object is a tiger and has four legs allows us to draw exactly the same (nonmonotonic) inferences as knowing only that it is a tiger.

What cumulative gives us is that the prototypical region stays the same if the new information we learn is already valid of the objects of the original prototypical region. But what happens if $\alpha \not\sim \beta$? This is where the structure of concepts and prototypes comes in. In the general case, learning more about a description means revising the concept we had in mind.

We use the fallbacks around the concept and around the prototypical region to generate the new concept and the new prototypical region:

Definition 5.4 Let α be a description and $p([\alpha])$ the set of prototypical α -objects. After learning that the object described satisfies property β , the new description is given by $\alpha * \beta$, with $[\alpha * \beta] = [\beta] \cap c_\alpha(\beta)$, where $c_\alpha(\beta)$ is the minimal fallback around $[\alpha]$ that intersects $[\beta]$. The new prototypical region is given by $p'([\alpha * \beta]) = [\beta] \cap c_{p\alpha}(\beta)$, where $c_{p\alpha}(\beta)$ is the minimal fallback around $p([\alpha])$ that intersects $[\beta]$.

From the cumulativity result, it follows that if $\alpha \sim \beta$, then $p'([\alpha * \beta]) = p([\alpha \wedge \beta]) = p([\alpha])$.

6 Conclusions and Further Work

We have extended the work of [Leh98] in order to deal with cases where new information about an object is learnt that is not compatible with the information we already had. This is the counterpart of an operation of belief revision, where one has to give up some information to accommodate new data but want to preserve as much information as possible. We do this by applying Grove's sphere model to sets of possible objects.

¹This is a combination of the properties called Aizerman and Cut in [Lin91].

The assumption that the properties that describe a concept can be totally ordered so that the weakenings of the description are nested is a bit unrealistic. An idea for future work is to use non-nested fallbacks. Rabinovicz and Lindström [LR91] present a modeling of relational belief revision in terms of non-nested fallbacks. The fallbacks are then “ellipsoids” containing the original theory K . In this way we can account for the fact that some of the dimensions of a concept are independent from each other.

It is well known that the AGM postulates correspond to properties of nonmonotonic inference [MG91]. It would be interesting to study the counterparts of the postulates in the context of concept revision.

Acknowledgments: I would like to thank Carlos Areces, Peter Gärdenfors and Frans Voorbraak for comments on an early draft. I am supported by the Brazilian national funding agency CAPES. This work was started during a visit to Lund University Cognitive Science, sponsored by the Dutch Research Council (NWO).

References

- [Gär88] Peter Gärdenfors. *Knowledge in Flux - Modeling the Dynamics of Epistemic States*. MIT Press, 1988.
- [Gär98] Peter Gärdenfors. *Conceptual Spaces*. book manuscript - to appear, 1998.
- [Gab85] Dov Gabbay. Theoretical foundations for nonmonotonic reasoning in expert systems. In K. Apt, editor, *Expert Systems, Logics and Models of Concurrent Systems*, pages 439–459, 1985.
- [GR95] Peter Gärdenfors and Hans Rott. Belief revision. In *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume IV, chapter 4.2. 1995.
- [Gro88] Adam Grove. Two modellings for theory change. *Journal of Philosophical Logic*, 17:157–170, 1988.
- [Han98] Sven Ove Hansson. *A Textbook of Belief Dynamics*. Kluwer Academic Press, 1998.
- [KLM90] Sarit Kraus, Daniel Lehmann, and Menachem Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.
- [Lak87] George Lakoff. *Women, Fire and Dangerous Things*. The University of Chicago Press, 1987.
- [Leh98] Daniel Lehmann. Stereotypical reasoning: Logical properties. *Journal of the IGPL*, 6(1):49–58, 1998.
- [Lin91] Sten Lindström. A semantic approach to nonmonotonic reasoning: Inference operations and choice. (Unpublished Manuscript), 1991.
- [LR91] Sten Lindström and Włodzimierz Rabinowicz. Epistemic entrenchment with incomparabilities and relational belief revision. In André Fuhrmann and Michael Morreau, editors, *The Logic of Theory Change*, volume 465 of *Lecture Notes in Artificial Intelligence*. Springer-Verlag, 1991.
- [MG91] David Makinson and Peter Gärdenfors. Relations between the logic of theory change and nonmonotonic logic. In André Fuhrmann and Michael Morreau, editors, *The Logic of Theory Change*, volume 465 of *Lecture Notes in Artificial Intelligence*. Springer-Verlag, 1991.
- [Ros73] Eleanor H. Rosch. Natural categories. *Cognitive Psychology*, 4:328–350, 1973.