



**HAL**  
open science

# A pre-semantics for counterfactual conditionals and similar logics

Karl Schlechta

► **To cite this version:**

Karl Schlechta. A pre-semantics for counterfactual conditionals and similar logics. 2016. hal-01408541v1

**HAL Id: hal-01408541**

**<https://hal.science/hal-01408541v1>**

Submitted on 5 Dec 2016 (v1), last revised 4 Apr 2017 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A pre-semantics for counterfactual conditionals and similar logics

Karl Schlechta <sup>\*†</sup>

December 5, 2016

## Abstract

The elegant Stalnaker/Lewis semantics for counterfactual conditionals works with distances between models. But human beings certainly have no tables of models and distances in their head.

We begin here an investigation using a more realistic picture, based on findings in neuroscience. We call it a pre-semantics, as its meaning is not a description of the world, but of the brain, whose structure is (partly) determined by the world it reasons about.

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Components</b>	<b>1</b>
2.1	Areas . . . . .	1
2.2	Connections between areas . . . . .	2
2.3	Attention . . . . .	2
2.4	Composing areas . . . . .	2
2.5	Majority rule . . . . .	3
<b>3</b>	<b>Comparisons and properties</b>	<b>3</b>
3.1	Pre-semantics and semantics . . . . .	3
3.2	Areas and models or model sets . . . . .	3
3.3	Inferential distance . . . . .	3
3.4	The Stalnaker/Lewis semantics for counterfactual conditionals . . . . .	4
3.5	Preferences and the hierarchy in [GS16] . . . . .	4
3.6	Defeasible inheritance diagrams . . . . .	4
3.7	Reactive arrows . . . . .	4
<b>4</b>	<b>Shortcomings</b>	<b>4</b>
	<b>References</b>	<b>6</b>

---

\*schcsg@gmail.com - <https://sites.google.com/site/schlechtakarl/> - Koppeweg 24, D-97833 Frammersbach, Germany

†Retired, formerly: Aix-Marseille Université, CNRS, LIF UMR 7279, F-13000 Marseille, France

# 1 Introduction

The idea is in its initial stages, and much work needs to be done to flesh it out. We are not even sure if we found the right level of abstraction, time and experience will tell. Still, we think the basic idea is promising, and would like to present it for discussion.

We have three types of objects:

- (1) areas,
- (2) connections,
- (3) attention.

Areas will be connected areas of the brain, corresponding to some information, areas may connect to other areas with different types of connectors, which may be positive (attractions) or negative (repulsions). Finally, attention focusses on areas or parts of areas, and their connections, or part thereof.

Areas do not necessarily correspond to nodes of graphs, as they are not atomic, and they can combine to new areas. Attention may hide contradictions (repulsions), so the whole picture may be contradiction tolerant.

Neurons have a direction, so will connections.

The brain usually is strongly redundant, democracy of areas and connections will provide robustness.

## 2 Components

### 2.1 Areas

We use the word “area” to designate

- on the physiological level a (perhaps only momentarily) strongly connected area of the brain,
- on the meaning level a picture, scene (in the sense of conscious scene), a prototype (without all the connotations the word “prototype” might have), any fragment of information. It need not be complete with all important properties, birds which fly, etc., it might be a robin sitting on a branch in sunshine, just any bit of information, abstract, concrete, mixture of both, whatever.

Thus, we are deliberately vague.

In particular, an area will not necessarily be atomic, we leave the possibility of looking inside, splitting an area into several areas, open. Areas need not be pre-determined, they may be formed and dissolved dynamically. (This will not be discussed here.)

### 2.2 Connections between areas

Areas can connect to each other. We simplify, as we separate surface from interior of areas. The picture we have in mind is that of a cell or a virus, which present “docking possibilities” on their surface. But, again, real areas need not have a surface, separate from the interior, through which all outside contacts have to pass.

We simplify, the “docking parts” will have two parts, a quality, and a value. Quality may, for instance, be colour, species, etc. Value may then be black, raven, car, etc. For simplicity, qualities will correspond to expressions in a suitable language. This makes manipulation easier. We suppose that only connectors of the same quality can connect to each other. Colour can connect to colour, but not colour to species. (A black raven is an area, black and raven may be connected inside the area. It is a simplification!) In analogy to nerves, we suppose that connections have a direction. If we have connections in both directions, we can always work with a (coupled) pair.

A connection between two connectors can be variously strong. For instance, the connection between  $\langle colour, black \rangle$  and  $\langle colour, grey \rangle$  will be stronger than that between  $\langle colour, black \rangle$  and  $\langle colour, red \rangle$ . If the

values differ too much, then it is rather a repulsion, than a poor attraction. Obviously, having no connection of a certain type is different from having a repulsion of this type.

Two (or more) areas may connect via more than one connector, e.g. the area “raven” and the area “blackbird” may connect via “bird” and via “colour”. Connected areas may sometimes be seen as one area, which can connect to new areas or other composite areas. This process is not necessarily commutative, etc., see below. Two areas may have several connections for the same quality. The overall strength will then be (somehow) calculated, adding (and perhaps subtracting) the individual strengths.

## 2.3 Attention

A third ingredient is “attention”. We picture attention as a light which shines on some areas, perhaps only on parts of those areas, and their connections, or only parts of the connections.

Attention allows, among other things, to construct a coherent picture by focussing only on parts of the picture, which are coherent. In particular, we might focus our attention on coherences, e.g., when we want consolidate a theory, or incoherences, when we want to attack a theory. Focussing on coherences might hide serious flaws in a theory, or our thinking in general. In context  $A$ , we might focus on  $\alpha$ , in context  $B$ , on  $\beta$ , etc.

As we leave attention deliberately unregulated, changes in attention may have very “wild” consequences.

## 2.4 Composing areas

Areas can be composed.

We may compose “black movement in the sky” via “black”, “flies” with “raven” to the picture “black raven flies overhead” or so.

Not all connectible areas need to be composed. Either they (or their connections) are not in the focus of attention, or the connections are considered (presently!) too weak for composition (perhaps in competition with other possible connections). The coherence of the composed area is the strength of the connections (for simplicity: of the inter-area connections), minus the repulsions.

The robustness is measured under varying attention. We may focus on repulsions, or just different connectors.

Suppose we try to compose area  $A$  with areas of type  $\beta$  and  $\gamma$ . When choosing area  $B$  of type  $\beta$ , we might choose by the strength of the connections between  $A$  and  $B$ . Choosing now  $C$  of type  $\gamma$ , we may choose by the strength of the connections between composed area  $A + B$  and  $C$ . Of course, the order may matter. Choosing first area  $C'$  of type  $\gamma$  to compose with  $A$ , and then  $B'$  to compose with  $A + C$  need not result in  $B = B'$  and  $C = C'$ . Choosing simultaneously  $B''$  and  $C''$  to give the best overall result  $A + B'' + C''$  might give still different  $B''$  and  $C''$ .

## 2.5 Majority rule

It was Edelman’s insight, see e.g. [Ede89], [Ede04], to see the parallels between the brain and the immune system, both working with selection from many possibilities. We imitate this principle. When areas are combined, we evaluate the results by its coherence, the strengths of the connections between the elements. We assume that we have many candidates of the same type, so we have a population from which to choose. We choose the best (by coherence), and consider this set for the properties of those combined areas. This is a weighed democracy, and offers some robustness.

# 3 Comparisons and properties

As we left the attention deliberately unregulated, it seems very difficult to obtain any non-trivial properties under changing attention.

But even with fixed attention, it seems difficult to obtain the usual logical properties like AND, OR, CUM, etc.

### 3.1 Pre-semantics and semantics

There is a conceptual difference between “real” semantics, and our pre-semantics.

Real semantics interpret language and logic in (an abstraction of) the world. Pre-semantics is an abstraction of (the functioning of) the brain. Of course, the brain is “somehow” connected to the world, but this would then be a semantics of (the functioning of) the brain. Thus, this pre-semantics is an intermediate step between language and the world.

### 3.2 Areas and models or model sets

Areas do NOT necessarily correspond to models or model sets, be they partial or not. For instance, the area “bird” need not be connected or strongly connected to all birds, species of birds, etc., which we know. The connection might be too weak, rarely considered, or not in the focus of our attention. In classical logic, there is no cost involved for going from model sets to subsets, or individual models, and back. Depending on attention and strength of connections, the cost might here be considerable.

Thus, a semantics for counterfactual conditionals based on this idea of areas will be different from the usual semantics for counterfactual conditionals, which is monotone in the first argument. (The more general the present situation is, the more “closest” situations will be considered.)

### 3.3 Inferential distance

In a composed area, we have an “inferential distance”, say from subarea  $a$  to subarea  $c$  via subarea  $b$ , perhaps over various connections. This is then a “natural” inferential distance and not one determined by an “artificial” proof schema. Likewise, we can complete a definition, say of a certain animal, when necessary, using more “far fetched” properties.

### 3.4 The Stalnaker/Lewis semantics for counterfactual conditionals

To see what happens with counterfactual conditionals, we consider the following example: “If it were to rain, I would take an umbrella, despite the strong wind.”

First, we have the present situation. Then, we have the picture of an umbrella protecting against the rain. We also have the picture of an umbrella which is blown inside out in a strong wind. Then we have pictures of situations where we would not take an umbrella, wind or not, e.g., we have to use both hands to carry our Christmas presents, and no way to carry an (open) umbrella. Then, we have situations where an umbrella is useful even in strong wind, e.g., we just come from an expensive hairdresser, and protecting our hairdo is all we need, and we have both hands free.

To construct a picture corresponding to the situation and its implications, we need to consider pictures of rain and umbrella, wind and umbrella, reasons why we use an umbrella, reasons why we might even use an umbrella in strong wind, etc. There is no ready picture available, but we have to combine several pictures, and make them fit with the original situation.

The minimal change idea of the Stalnaker/Lewis semantics, see e.g. [Sta68], [Lew73], is replaced here by coherence. The more things we have to modify, the less good coherence will be.

We have to separate looking for pictures, and combining them. Language helps in both tasks, addressing and combining areas liberally.

Note that the case of update is somewhat different. In update, the question is how the actual world changes, and not how we believe the world (and we) behave, as is the case here.

### 3.5 Preferences and the hierarchy in [GS16]

It is natural to combine the ideas of the hierarchy in [GS16], chapter 11 there, with our present ideas. Exceptional classes, like penguins, are only loosely bound to regular classes, like birds; surprise cases even more loosely.

### 3.6 Defeasible inheritance diagrams

Inheritance diagrams give in general different results for upward and downward chaining. In this sense, they are similar to our approach (first combine  $a$  with  $b$ , then with  $c$ , or first with  $c$ , then with  $b$ ). Length of path approaches are somewhat “global”, as determining the quality of a composed area by a global assessment here, is. Thus, we find some similar considerations in the theory of defeasible inheritance networks.

For an overview of defeasible inheritance, see e.g. [Sch97-2].

### 3.7 Reactive arrows

As we can have several connections of the same quality, there is a similarity to Gabbay’s reactive arrows, see e.g. [Gab04], [GS08e]: We can cancel one connection by another one, which has opposite polarity and is stronger.

## 4 Shortcomings

In our examples, we only mentioned static pictures, but there is no restriction - we can also work with dynamic pictures, birds, which do not fly at the moment, but are able to fly, etc.

We can combine areas, but our ability to look inside areas is limited. The only way to do this in our outline is by focussing our attention. In a further development, one might want to break up existing areas, or, conversely, make internal connections stronger.

One might also like to have an automatic degradation of connections which were not used recently.

In this first draft, we treat the areas we find basically as atoms, though we know that they are not. It is a simplification. In addition, in the “real life of the brain” there is not necessarily a surface vs. an interior of the areas, all this is dynamical. The best definition might be coherence of areas by number and strength of the interior connections vs. those of the exterior of the area. For instance, it might be interesting to separate acoustical from visual properties.

Negation can be introduced by an inversion of values,  $\langle colour, black \rangle$  has then to pass through  $\langle colour, not - black \rangle$ . We do not know if this is realistic.

## References

- [Ede04] Gerald M. Edelman, “Wider than the sky”, Yale University Press, New Haven 2004, (German edition “Das Licht des Geistes”, Rowohlt, 2007)
- [Ede89] Gerald M. Edelman, “The remembered present”, Basic Books, New York, 1989
- [GS08e] D. Gabbay, K. Schlechta, “Defeasible inheritance systems and reactive diagrams”, *Logic Journal of the IGPL*, 17:1–54, 2009
- [GS16] D. Gabbay, K. Schlechta, “A New Perspective on Nonmonotonic Logics”, Springer, Heidelberg, Nov. 2016, ISBN 978-3-319-46815-0,
- [Gab04] D. M. Gabbay, “Reactive Kripke semantics and arc accessibility”, *Proceedings CombLog04*, July 28–30, 2004, W. Carnielli, F. M. Dionesio, P. Mateus, Eds., Centre of Logic and Computation, University of Lisbon 2004, pp. 7–20.
- [Lew73] D. Lewis, “Counterfactuals”, Blackwell, Oxford, 1973
- [Sch97-2] K. Schlechta, “Nonmonotonic logics - basic concepts, results, and techniques” Springer Lecture Notes series, LNAI 1187, p.243, Jan. 1997.
- [Sta68] R. Stalnaker, “A theory of conditionals”, N. Rescher (ed.), “Studies in logical theory”, Blackwell, Oxford, pp. 98–112