



HAL
open science

A semantics for obligations

Dov Gabbay, Karl Schlechta

► **To cite this version:**

| Dov Gabbay, Karl Schlechta. A semantics for obligations. 2008. hal-00339393v2

HAL Id: hal-00339393

<https://hal.science/hal-00339393v2>

Preprint submitted on 9 Mar 2009

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 semantics for obligations - Local and global properties of obligations

*

Dov M Gabbay [†]
King's College, London [‡]

Karl Schlechta [§]
Laboratoire d'Informatique Fondamentale de Marseille [¶]

March 9, 2009

Abstract

We analyze a number of properties obligations have or should have. We will not present a definite answer what an obligation “is”, but rather point out the numerous possibilities to consider, as well as some of their interrelations.

Contents

1	Introductory remarks	2
1.1	Context	2
1.2	Central idea	2
1.3	A common property of facts and obligations	2
1.4	Derivations of obligations	3
1.5	Orderings and obligations	3
1.6	Derivation revisited	3
1.7	Relativization	3
1.8	Numerous possibilities	3
1.9	Notation	4
1.10	Overview of the article	4
2	Basic definitions	4
2.1	Size	4
2.2	Distance	5
2.3	Quality and closure	6
2.4	Neighbourhood	9
2.5	Unions of intersections and other definitions	9
3	Philosophical discussion of obligations	10
3.1	A fundamental difference between facts and obligations: asymmetry and negation	10
3.2	“And” and “or” for obligations	10
3.3	Ceteris paribus - a local property	11
3.4	Hamming neighbourhoods	11
3.5	Global and mixed global/local properties of obligations	11
3.6	Soft obligations	12
3.7	Overview of different types of obligations	12
3.8	Summary of the philosophical remarks	13
4	Examination of the various cases	13
4.1	Hard obligations for the set approach	13

*Paper 332

[†]Dov.Gabbay@kcl.ac.uk, www.dcs.kcl.ac.uk/staff/dg

[‡]Department of Computer Science, King's College London, Strand, London WC2R 2LS, UK

[§]ks@cmi.univ-mrs.fr, karl.schlechta@web.de, http://www.cmi.univ-mrs.fr/~ks

[¶]UMR 6166, CNRS and Université de Provence, Address: CMI, 39, rue Joliot-Curie, F-13453 Marseille Cedex 13, France

4.1.1	Introduction	13
4.1.2	The not necessarily independent case	14
4.1.3	The independent case	15
4.2	Remarks on the counting case	16
5	What is an obligation?	16
6	Conclusion	17
7	Acknowledgements	17
	References	17

1 Introductory remarks

We see some relation of “better” as central for obligations. Obligations determine what is “better” and what is “worse”, conversely, given such a relation of “better”, we can define obligations. The problems lie, in our opinion, in the fact that an adequate treatment of such a relation is somewhat complicated, and leads to many ramifications.

On the other hand, obligations have sufficiently many things in common with facts so we can in a useful way say that an obligation is satisfied in a situation, and one can also define a notion of derivation for obligations.

Our approach is almost exclusively semantical.

1.1 Context

The problem with formalisation using logic is that the natural movements in the application area being formalised do not exactly correspond to natural movements in the logic being used as a tool of formalisation. Put differently, we may be able to express statement A of the application area by a formula ϕ of the logic, but the subtleties of the way A is manipulated in the application area cannot be matched in the formal logic used. This gives rise to paradoxes. To resolve the paradoxes one needs to improve the logic. So the progress in the formalisation program depends on the state of development of logic itself. Recent serious advances in logical tools by the authors of this paper enable us to offer some better formalisations possibilities for the notion of obligation. This is what we offer in this paper.

Historically, articles on Deontic Logic include collections of problems, see e.g. [MDW94], semantical approaches, see e.g. [Han69], and others, like [CJ02].

Our basic idea is to see obligations as tightly connected to some relation of “better”. An immediate consequence is that negation, which inverses such a relation, behaves differently in the case of obligations and of classical logic. (“And” and “or” seem to show analogue behaviour in both logics.) The relation of “better” has to be treated with some caution, however, and we introduce and investigate local and global properties about “better” of obligations. Most of these properties coincide in sufficiently nice situations, in others, they are different.

We do not come to a final conclusion which properties obligations should or should not have, perhaps this will be answered in future, perhaps there is no universal answer. We provide a list of ideas which seem reasonable to us.

Throughout, we work in a finite (propositional) setting.

1.2 Central idea

We see a tight connection between obligations and a relation of “morally” better between situations. Obligations are there to guide us for “better” actions, and, conversely, given some relation of “better”, we can define obligations.

The problems lie in the fact that a simple approach via quality is not satisfactory. We examine a number of somewhat more subtle ideas, some of them also using a notion of distance.

1.3 A common property of facts and obligations

We are fully aware that an obligation has a conceptually different status than a fact. The latter is true or false, an obligation has been set by some instance as a standard for behaviour, or whatever.

Still, we will say that an obligation holds in a situation, or that a situation satisfies an obligation. If the letter is in the mail box, the obligation to post it is satisfied, if it is in the trash bin, the obligation is not satisfied. In some set of worlds, the obligation is satisfied, in the complement of this set, it is not satisfied. Thus, obligations behave in this respect like facts, and we put for this aspect all distinctions between facts and obligations aside.

Thus, we will treat obligations most of the time as subsets of the set of all models, but also sometimes as formulas. As we work mostly in a finite propositional setting, both are interchangeable.

We are *not* concerned here with actions to fulfill obligations, developments or so, just simply situations which satisfy or not obligations.

This is perhaps also the right place to point out that one has to distinguish between facts that *hold* in “good” situations (they will be closed under arbitrary right weakening), and obligations which describe what *should* be, they will not be closed under arbitrary right weakening. This article is only about the latter.

1.4 Derivations of obligations

Again, we are aware that “deriving” obligations is different from “deriving” facts. Derivation of facts is supposed to conclude from truth to truth, deriving obligations will be about concluding what can also be considered an obligation, given some set of “basic” obligations. The parallel is sufficiently strong to justify the double use of the word “derive”.

Very roughly, we will say that conjunctions (or intersections) and disjunctions (unions) of obligations lead in a reasonable way to derived obligations, but negations do not. We take the Ross paradox (see below) very seriously, it was, as a matter of fact, our starting point to avoid it in a reasonable notion of derivation.

We mention two simple postulates derived obligations should probably satisfy.

- (1) Every original obligation should also be a derived obligation, corresponding to $\alpha, \beta \sim \alpha$.
- (2) A derived obligation should not be trivial, i.e. neither empty nor U , the universe we work in.

The last property is not very important from an algebraic point of view, and easily satisfiable, so we will not give it too much attention.

1.5 Orderings and obligations

There is, in our opinion, a deep connection between obligations and orderings (and, in addition, distances), which works both ways.

First, given a set of obligations, we can say that one situation is “better” than a second situation, if the first satisfies “more” obligations than the second does. “More” can be measured by the set of obligations satisfied, and then by the subset/superset relation, or by the number of obligations. Both are variants of the Hamming distance. “Distance” between two situations can be measured by the set or number of obligations in which they differ (i.e. one situation satisfies them, the other not). In both cases, we will call the variants the set or the counting variant.

This is also the deeper reason why we have to be careful with negation. Negation inverses such orderings, if ϕ is better than $\neg\phi$, then $\neg\phi$ is worse than $\neg\neg\phi = \phi$. But in some reasonable sense \wedge and \vee preserve the ordering, thus they are compatible with obligations.

Conversely, given a relation (of quality), we might for instance require that obligations are closed under improvement. More subtle requirements might work with distances. The relations of quality and distance can be given abstractly (as the notion of size used for “soft” obligations), or as above by a starting set of obligations. We will also define important auxiliary concepts on such abstract relations.

1.6 Derivation revisited

A set of “basic” obligations generates an ordering and a distance between situations, ordering and distance can be used to define properties obligations should have. It is thus natural to define obligations derived from the basic set as those sets of situations which satisfy the desirable properties of obligations defined via the order and distance generated by the basic set of obligations. Our derivation is thus a two step procedure: first generate the order and distance, which define suitable sets of situations.

We will call properties which are defined without using distances global properties (like closure under improving quality), properties involving distance (like being a neighbourhood) local properties.

1.7 Relativization

An important problem is relativization. Suppose \mathcal{O} is a set of obligations for all possible situations, e.g. O is the obligation to post the letter, and O' is the obligation to water the flowers. Ideally, we will do both. Suppose we consider now a subset, where we cannot do both (e.g. for lack of time). What are our obligations in this subset? Are they just the restrictions to the subset? Conversely, if O is an obligation for a subset of all situations, is then some O' with $O \subseteq O'$ an obligation for the set of all situations?

In more complicated requirements, it might be reasonable e.g. to choose the ideal situations still in the big set, even if they are not in the subset to be considered, but use an approximation inside the subset. Thus, relativizations present a non-trivial problem with many possible solutions, and it seems doubtful whether a universally adequate solution can be found.

1.8 Numerous possibilities

Seeing the possibilities presented so far (set or counting order, set or counting distance, various relativizations), we can already guess that there are numerous possible reasonable approaches to what an obligation is or should be. Consequently, it seems quite impossible to pursue all these combinations in detail. Thus, we concentrate mostly on one combination, and leave it to the reader to fill in details for the others, if (s)he is so interested.

We will also treat the defeasible case here. Perhaps somewhat surprisingly, this is straightforward, and largely due to the fact that there is one natural definition of “big” sets for the product set, given that “big” sets are defined for the components. So there are no new possibilities to deal with here.

The multitude of possibilities is somewhat disappointing. It may, of course, be due to an incapacity of the present authors to find *the* right notion. But it may also be a genuine problem of ambiguous intuitions, and thus generate conflicting views and misunderstandings on the one side, and loopholes for not quite honest argumentation in practical juridical reasoning on the other hand.

1.9 Notation

$\mathcal{P}(X)$ will be the powerset of X , $A \subseteq B$ will mean that A is a subset of B , or $A = B$, $A \subset B$ that A is a proper subset of B .

1.10 Overview of the article

We will work in a finite propositional setting, so there is a trivial and 1-1 correspondence between formulas and model sets. Thus, we can just work with model sets - which implies, of course, that obligations will be robust under logical reformulation. So we will formulate most results only for sets.

- In Section 2 (page 4) , we give the basic definitions, together with some simple results about those definitions.
- Section 3 (page 10) will present a more philosophical discussion, with more examples, and we will argue that our definitions are relevant for our purpose.
- As said already, there seems to be a multitude of possible and reasonable definitions of what an obligation can or should be, so we limit our formal investigation to a few cases, this is given in Section 4 (page 13).
- In Section 5 (page 16), we give a tentative definition of an obligation.

(The concept of neighbourhood semantics is not new, and was already introduced by D.Scott, [Sco70], and R.Montague, [Mon70]. Further investigations showed that it was also used by O.Pacheco, [Pac07], precisely to avoid unwanted weakening for obligations. We came the other way and started with the concept of independent strengthening, see Definition 2.13 (page 9), and introduced the abstract concept of neighbourhood semantics only at the end. This is one of the reasons we also have different descriptions which turned out to be equivalent: we came from elsewhere.)

2 Basic definitions

We give here all definitions needed for our treatment of obligations. The reader may continue immediately to Section 3 (page 10), and come back to the present section whenever necessary.

Intuitively, U is supposed to be the set of models of some propositional language \mathcal{L} , but we will stay purely algebraic whenever possible. $U' \subseteq U$ is some subset.

For ease of writing, we will here and later sometimes work with propositional variables, and also identify models with the formula describing them, still this is just shorthand for arbitrary sets and elements. pq will stand for $p \wedge q$, etc.

If a set \mathcal{O} of obligations is given, these will be just arbitrary subsets of the universe U . We will also say that \mathcal{O} is over U . Before we deepen the discussion of more conceptual aspects, we give some basic definitions (for which we claim no originality). We will need them quite early, and think it is better to put them together here, not to be read immediately, but so the reader can leaf back and find them easily.

We work here with a notion of size (for the defeasible case), a notion d of distance, and a quality relation \preceq . The latter two can be given abstractly, but may also be defined from a set of (basic) obligations.

We use these notions to describe properties obligations should, in our opinion, have. A careful analysis will show later interdependencies between the different properties.

2.1 Size

For each $U' \subseteq U$ we suppose an abstract notion of size to be given. We may assume this notion to be a filter or an ideal. Coherence properties between the filters/ideals of different U', U'' will be left open, the reader may assume them to be the conditions of the system P of preferential logic, see [GS09a].

Given such notions of size on U' and U'' , we will also need a notion of size on $U' \times U''$. We take the evident solution:

Definition 2.1

Let a notion of “big subset” be defined by a principal filter for all $X \subseteq U$ and all $X' \subseteq U'$. Thus, for all $X \subseteq U$ there exists a fixed principal filter $\mathcal{F}(X) \subseteq \mathcal{P}(X)$, and likewise for all $X' \subseteq U'$. (This is the situation in the case of preferential structures, where $\mathcal{F}(X)$ is generated by $\mu(X)$, the set of minimal elements of X .)

Define now $\mathcal{F}(X \times X')$ as generated by $\{A \times A' : A \in \mathcal{F}(X), A' \in \mathcal{F}(X')\}$, i.e. if A is the smallest element of $\mathcal{F}(X)$, A' the smallest element of $\mathcal{F}(X')$, then $\mathcal{F}(X \times X') := \{B \subseteq X \times X' : A \times A' \subseteq B\}$.

Fact 2.1

If $\mathcal{F}(X)$ and $\mathcal{F}(X')$ are generated by preferential structures $\preceq_X, \preceq_{X'}$, then $\mathcal{F}(X \times X')$ is generated by the product structure defined by

$$\langle x, x' \rangle \preceq_{X \times X'} \langle y, y' \rangle \Leftrightarrow x \preceq_X y \text{ and } x' \preceq_{X'} y'.$$

Proof

We will omit the indices of the orderings when this causes no confusion.

Let $A \in \mathcal{F}(X)$, $A' \in \mathcal{F}(X')$, i.e. A minimizes X , A' minimizes X' . Let $\langle x, x' \rangle \in X \times X'$, then there are $a \in A$, $a' \in A'$ with $a \preceq x$, $a' \preceq x'$, so $\langle a, a' \rangle \preceq \langle x, x' \rangle$.

Conversely, suppose $U \subseteq X \times X'$, U minimizes $X \times X'$, but there is no $A \times A' \subseteq U$ s.t. $A \in \mathcal{F}(X)$, $A' \in \mathcal{F}(X')$. Assume $A = \mu(X)$, $A' = \mu(X')$, so there is $\langle a, a' \rangle \in \mu(X) \times \mu(X')$, $\langle a, a' \rangle \notin U$. But only $\langle a, a' \rangle \preceq \langle a, a' \rangle$, and U does not minimize $X \times X'$, *contradiction*.

□

Note that a natural modification of our definition:

There is $A \in \mathcal{F}(X)$ s.t. for all $a \in A$ there is a (maybe varying) $A'_a \in \mathcal{F}(X')$, and $U := \{\langle a, a' \rangle : a \in A, a' \in A'_a\}$ as generating sets

will result in the same definition, as our filters are principal, and thus stable under arbitrary intersections.

2.2 Distance

We consider a set of sequences Σ , for $x \in \Sigma$ $x : I \rightarrow S$, I a finite index set, S some set. Often, S will be $\{0, 1\}$, $x(i) = 1$ will mean that $x \in i$, when $I \subseteq \mathcal{P}(U)$ and $x \in U$. For abbreviation, we will call this (unsystematically, often context will tell) the \in -case. Often, I will be written \mathcal{O} , intuitively, $O \in \mathcal{O}$ is then an obligation, and $x(O) = 1$ means $x \in O$, or x “satisfies” the obligation O .

Definition 2.2

In the \in -case, set $\mathcal{O}(x) := \{O \in \mathcal{O} : x \in O\}$.

Definition 2.3

Given $x, y \in \Sigma$, the Hamming distance comes in two flavours:

$d_s(x, y) := \{i \in I : x(i) \neq y(i)\}$, the set variant,

$d_c(x, y) := \text{card}(d_s(x, y))$, the counting variant.

We define $d_s(x, y) \leq d_s(x', y')$ iff $d_s(x, y) \subseteq d_s(x', y')$,

thus, s -distances are not always comparable.

Fact 2.2

(1) In the \in -case, we have $d_s(x, y) = \mathcal{O}(x) \Delta \mathcal{O}(y)$, where Δ is the symmetric set difference.

(2) d_c has the normal addition, set union takes the role of addition for d_s , \emptyset takes the role of 0 for d_s , both are distances in the following sense:

(2.1) $d(x, y) = 0$ if $x = y$, but not conversely,

(2.2) $d(x, y) = d(y, x)$,

(2.3) the triangle inequality holds, for the set variant in the form $d_s(x, z) \subseteq d_s(x, y) \cup d_s(y, z)$.

(If $d(x, y) = 0 \not\Rightarrow x = y$ poses a problem, one can always consider equivalence classes.)

Proof

(2.1) Suppose $U = \{x, y\}$, $\mathcal{O} = \{U\}$, then $\mathcal{O}(X) = \mathcal{O}(Y)$, but $x \neq y$.

(2.3) If $i \notin d_s(x, y) \cup d_s(y, z)$, then $x(i) = y(i) = z(i)$, so $x(i) = z(i)$ and $i \notin d_s(x, z)$.

The others are trivial.

□

Definition 2.4

(1) We can define for any distance d with some minimal requirements a notion of “between”.

If the codomain of d has an ordering \leq , but no addition, we define:

$\langle x, y, z \rangle_d : \Leftrightarrow d(x, y) \leq d(x, z)$ and $d(y, z) \leq d(x, z)$.

If the codomain has a commutative addition, we define

$\langle x, y, z \rangle_d : \Leftrightarrow d(x, z) = d(x, y) + d(y, z)$ - in $d_s +$ will be replaced by \cup , i.e.

$\langle x, y, z \rangle_s : \Leftrightarrow d(x, z) = d(x, y) \cup d(y, z)$.

For above two Hamming distances, we will write $\langle x, y, z \rangle_s$ and $\langle x, y, z \rangle_c$.

(2) We further define:

$[x, z]_d := \{y \in X : \langle x, y, x \rangle_d\}$ - where X is the set we work in.

We will write $[x, z]_s$ and $[x, z]_c$ when appropriate.

(3) For $x \in U$, $X \subseteq U$ set $x \parallel_d X := \{x' \in X : \neg \exists x'' \neq x' \in X. d(x, x'') \geq d(x, x')\}$.

Note that, if $X \neq \emptyset$, then $x \parallel X \neq \emptyset$.

We omit the index when this does not cause confusion. Again, when adequate, we write \parallel_s and \parallel_c .

For problems with characterizing “between” see [Sch04].

Fact 2.3

(0) $\langle x, y, z \rangle_d \Leftrightarrow \langle z, y, x \rangle_d$.

Consider the situation of a set of sequences Σ .

Let $A := A_{\sigma, \sigma''} := \{\sigma' : \forall i \in I (\sigma(i) = \sigma''(i) \rightarrow \sigma'(i) = \sigma(i) = \sigma''(i))\}$. Then

- (1) If $\sigma' \in A$, then $d_s(\sigma, \sigma'') = d_s(\sigma, \sigma') \cup d_s(\sigma', \sigma'')$, so $\langle \sigma, \sigma', \sigma'' \rangle_s$.
- (2) If $\sigma' \in A$ and S consists of 2 elements (as in classical 2-valued logic), then $d_s(\sigma, \sigma')$ and $d_s(\sigma', \sigma'')$ are disjoint.
- (3) $[\sigma, \sigma'']_s = A$.
- (4) If, in addition, S consists of 2 elements, then $[\sigma, \sigma'']_c = A$.

Proof

(0) Trivial.

(1) “ \subseteq ” follows from Fact 2.2 (page 5), (2.3).

Conversely, if e.g. $i \in d_s(\sigma, \sigma')$, then by prerequisite $i \in d_s(\sigma, \sigma'')$.

(2) Let $i \in d_s(\sigma, \sigma') \cap d_s(\sigma', \sigma'')$, then $\sigma(i) \neq \sigma'(i)$ and $\sigma'(i) \neq \sigma''(i)$, but then by $\text{card}(S) = 2$ $\sigma(i) = \sigma''(i)$, but $\sigma' \in A$, *contradiction*.

We turn to (3) and (4):

If $\sigma' \notin A$, then there is i' s.t. $\sigma(i') = \sigma''(i') \neq \sigma'(i')$. On the other hand, for all i s.t. $\sigma(i) \neq \sigma''(i)$ $i \in d_s(\sigma, \sigma') \cup d_s(\sigma', \sigma'')$. Thus:

- (3) By (1) $\sigma' \in A \Rightarrow \langle \sigma, \sigma', \sigma'' \rangle_s$. Suppose $\sigma' \notin A$, so there is i' s.t. $i' \in d_s(\sigma, \sigma') - d_s(\sigma, \sigma'')$, so $\langle \sigma, \sigma', \sigma'' \rangle_s$ cannot be.
- (4) By (1) and (2) $\sigma' \in A \Rightarrow \langle \sigma, \sigma', \sigma'' \rangle_c$. Conversely, if $\sigma' \notin A$, then $\text{card}(d_s(\sigma, \sigma')) + \text{card}(d_s(\sigma', \sigma'')) \geq \text{card}(d_s(\sigma, \sigma'')) + 2$.

□

Definition 2.5

Given a finite propositional language \mathcal{L} defined by the set $v(\mathcal{L})$ of propositional variables, let \mathcal{L}_\wedge be the set of all consistent conjunctions of elements from $v(\mathcal{L})$ or their negations. Thus, $p \wedge \neg q \in \mathcal{L}_\wedge$ if $p, q \in v(\mathcal{L})$, but $p \vee q, \neg(p \wedge q) \notin \mathcal{L}_\wedge$. Finally, let $\mathcal{L}_{\vee\wedge}$ be the set of all (finite) disjunctions of formulas from \mathcal{L}_\wedge . (As we will later not consider all formulas from \mathcal{L}_\wedge , this will be a real restriction.)

Given a set of models M for a finite language \mathcal{L} , define $\phi_M := \bigwedge \{p \in v(\mathcal{L}) : \forall m \in M. m(p) = v\} \wedge \bigwedge \{\neg p : p \in v(\mathcal{L}), \forall m \in M. m(p) = f\} \in \mathcal{L}_\wedge$. (If there are no such p , set $\phi_M := \text{TRUE}$.)

This is the strongest $\phi \in \mathcal{L}_\wedge$ which holds in M .

Fact 2.4

If x, y are models, then $[x, y] = M(\phi_{\{x, y\}})$. □

Proof

$m \in [x, y] \Leftrightarrow \forall p (x \models p, y \models p \Rightarrow m \models p \text{ and } x \not\models p, y \not\models p \Rightarrow m \not\models p), m \models \phi_{\{x, y\}} \Leftrightarrow m \models \bigwedge \{p : x(p) = y(p) = v\} \wedge \bigwedge \{\neg p : x(p) = y(p) = f\}$.

2.3 Quality and closure

Definition 2.6

Given any relation \preceq (of quality), we say that $X \subseteq U$ is (downward) closed (with respect to \preceq) iff $\forall x \in X \forall y \in U (y \preceq x \Rightarrow y \in X)$.

(Warning, we follow the preferential tradition, “smaller” will mean “better”.)

We think that being closed is a desirable property for obligations: what is at least as good as one element in the obligation should be “in”, too.

Fact 2.5

Let \preceq be given.

- (1) Let $D \subseteq U' \subseteq U''$, D closed in U'' , then D is also closed in U' .
- (2) Let $D \subseteq U' \subseteq U''$, D closed in U' , U' closed in U'' , then D is closed in U'' .
- (3) Let $D_i \subseteq U'$ be closed for all $i \in I$, then so are $\bigcup \{D_i : i \in I\}$ and $\bigcap \{D_i : i \in I\}$.

Proof

(1) Trivial.

(2) Let $x \in D \subseteq U'$, $x' \preceq x$, $x' \in U''$, then $x' \in U'$ by closure of U'' , so $x' \in D$ by closure of U' .

(3) Trivial.

□

We may have an abstract relation \preceq of quality on the domain, but we may also define it from the structure of the sequences, as we will do now.

Definition 2.7

Consider the case of sequences.

Given a relation \preceq (of quality) on the codomain, we extend this to sequences in Σ :

$$x \sim y :\Leftrightarrow \forall i \in I(x(i) \sim y(i))$$

$$x \preceq y :\Leftrightarrow \forall i \in I(x(i) \preceq y(i))$$

$$x \prec y :\Leftrightarrow \forall i \in I(x(i) \preceq y(i)) \text{ and } \exists i \in I(x(i) \prec y(i))$$

In the \in -case, we will consider $x \in i$ better than $x \notin i$. As we have only two values, true/false, it is easy to count the positive and negative cases (in more complicated situations, we might be able to multiply), so we have an analogue of the two Hamming distances, which we might call the Hamming quality relations.

Let \mathcal{O} be given now.

(Recall that we follow the preferential tradition, “smaller” will mean “better”.)

$$x \sim_s y :\Leftrightarrow \mathcal{O}(x) = \mathcal{O}(y),$$

$$x \preceq_s y :\Leftrightarrow \mathcal{O}(y) \subseteq \mathcal{O}(x),$$

$$x \prec_s y :\Leftrightarrow \mathcal{O}(y) \subset \mathcal{O}(x),$$

$$x \sim_c y :\Leftrightarrow \text{card}(\mathcal{O}(x)) = \text{card}(\mathcal{O}(y)),$$

$$x \preceq_c y :\Leftrightarrow \text{card}(\mathcal{O}(y)) \leq \text{card}(\mathcal{O}(x)),$$

$$x \prec_c y :\Leftrightarrow \text{card}(\mathcal{O}(y)) < \text{card}(\mathcal{O}(x)).$$

The requirement of closure causes a problem for the counting approach: Given e.g. two obligations O, O' , then any two elements in just one obligation have the same quality, so if one is in, the other should be, too. But this prevents now any of the original obligations to have the desirable property of closure. In the counting case, we will obtain a ranked structure, where elements satisfy 0, 1, 2, etc. obligations, and we are unable to differentiate inside those layers. Moreover, the set variant seems to be closer to logic, where we do not count the propositional variables which hold in a model, but consider them individually. For these reasons, we will not pursue the counting approach as systematically as the set approach. One should, however, keep in mind that the counting variant gives a ranking relation of quality, as all qualities are comparable, and the set variant does not. A ranking seems to be appreciated sometimes in the literature, though we are not really sure why.

Of particular interest is the combination of d_s and \preceq_s (d_c and \preceq_c) respectively - where by \preceq_s we also mean \prec_s and \sim_s , etc. We turn to this now.

Fact 2.6

We work in the \in -case.

$$(1) x \preceq_s y \Rightarrow d_s(x, y) = \mathcal{O}(x) - \mathcal{O}(y)$$

Let $a \prec_s b \prec_s c$. Then

(2) $d_s(a, b)$ and $d_s(b, c)$ are not comparable,

(3) $d_s(a, c) = d_s(a, b) \cup d_s(b, c)$, and thus $b \in [a, c]_s$.

This does not hold in the counting variant, as Example 2.1 (page 8) shows.

(4) Let $x \prec_s y$ and $x' \prec_s y$ with $x, x' \prec_s$ -incomparable. Then $d_s(x, y)$ and $d_s(x', y)$ are incomparable.

(This does not hold in the counting variant, as then all distances are comparable.)

(5) If $x \prec_s z$, then for all $y \in [x, z]_s$ $x \preceq_s y \preceq_s z$.

Proof

(1) Trivial.

(2) We have $\mathcal{O}(c) \subset \mathcal{O}(b) \subset \mathcal{O}(a)$, so the results follows from (1).(3) By definition of d_s and (1).(4) x and x' are \preceq_s -incomparable, so there are $O \in \mathcal{O}(x) - \mathcal{O}(x')$, $O' \in \mathcal{O}(x') - \mathcal{O}(x)$.

As $x, x' \prec_s y$, $O, O' \notin \mathcal{O}(y)$, so $O \in d_s(x, y) - d_s(x', y)$, $O' \in d_s(x', y) - d_s(x, y)$.

(5) $x \prec_s z \Rightarrow \mathcal{O}(z) \subset \mathcal{O}(x)$, $d_s(x, z) = \mathcal{O}(x) - \mathcal{O}(z)$. By prerequisite $d_s(x, z) = d_s(x, y) \cup d_s(y, z)$. Suppose $x \not\preceq_s y$. Then there is $i \in \mathcal{O}(y) - \mathcal{O}(x) \subseteq d_s(x, y)$, so $i \notin \mathcal{O}(x) - \mathcal{O}(z) = d_s(x, z)$, contradiction.

Suppose $y \not\prec_s z$. Then there is $i \in \mathcal{O}(z) - \mathcal{O}(y) \subseteq d_s(y, z)$, so $i \notin \mathcal{O}(x) - \mathcal{O}(z) = d_s(x, z)$, *contradiction*.

□

Example 2.1

In this and similar examples, we will use the model notation. Some propositional variables p, q , etc. are given, and models are described by $p \neg q r$, etc. Moreover, the propositional variables are the obligations, so in this example we have the obligations p, q, r .

Consider $x := \neg p \neg q r$, $y := p q \neg r$, $z := \neg p \neg q \neg r$. Then $y \prec_c x \prec_c z$, $d_c(x, y) = 3$, $d_c(x, z) = 1$, $d_c(z, y) = 2$, so $x \notin [y, z]_c$. □

Definition 2.8

Given a quality relation \prec between elements, and a distance d , we extend the quality relation to sets and define:

(1) $x \prec Y := \Leftrightarrow \forall y \in (x \parallel Y). x \prec y$. (The closest elements - i.e. there are no closer ones - of Y , seen from x , are less good than x .)

analogously $X \prec y := \Leftrightarrow \forall x \in (y \parallel X). x \prec y$

(2) $X \prec_l Y := \Leftrightarrow \forall x \in X. x \prec Y$ and $\forall y \in Y. X \prec y$ (X is locally better than Y).

When necessary, we will write $\prec_{l,s}$ or $\prec_{l,c}$ to distinguish the set from the counting variant.

For the next definition, we use the notion of size: $\nabla \phi$ iff for almost all ϕ holds i.e. the set of exceptions is small.

(3) $X \ll_l Y := \Leftrightarrow \nabla x \in X. x \prec Y$ and $\nabla y \in Y. X \prec y$.

We will likewise write $\ll_{l,s}$ etc.

This definition is supposed to capture quality difference under minimal change, the “ceteris paribus” idea: $X \prec_l CX$ should hold for an obligation X . Minimal change is coded by \parallel , and “ceteris paribus” by minimal change.

Fact 2.7

If $X \prec_l CX$, and $x \in U$ an optimal point (there is no better one), then $x \in X$.

Proof

If not, then take $x' \in X$ closest to x , this must be better than x , contradiction. □

Fact 2.8

Take the set version.

If $X \prec_{l,s} CX$, then X is downward \prec_s -closed.

Proof

Suppose $X \prec_{l,s} CX$, but X is not downward closed.

Case 1: There are $x \in X$, $y \notin X$, $y \sim_s x$. Then $y \in x \parallel_s CX$, but $x \not\prec_s y$, *contradiction*.

Case 2: There are $x \in X$, $y \notin X$, $y \prec_s x$. By $X \prec_{l,s} CX$, the elements in X closest to y must be better than y . Thus, there is $x' \prec_s y$, $x' \in X$, with minimal distance from y . But then $x' \prec_s y \prec_s x$, so $d_s(x', y)$ and $d_s(y, x)$ are incomparable by Fact 2.6 (page 7), so x is among those with minimal distance from y , so $X \prec_{l,s} CX$ does not hold. □

Example 2.2

We work with the set variant.

This example shows that \preceq_s -closed does not imply $X \prec_{l,s} CX$, even if X contains the best elements.

Let $\mathcal{O} := \{p, q, r, s\}$, $U' := \{x := p \neg q \neg r \neg s, y := \neg p q \neg r \neg s, x' := p q r s\}$, $X := \{x, x'\}$. x' is the best element of U' , so X contains the best elements, and X is downward closed in U' , as x and y are not comparable. $d_s(x, y) = \{p, q\}$, $d_s(x', y) = \{p, r, s\}$, so the distances from y are not comparable, so x is among the closest elements in X , seen from y , but $x \not\prec_s y$.

The lack of comparability is essential here, as the following Fact shows.

□

We have, however, for the counting variant:

Fact 2.9

Consider the counting variant. Then

If X is downward closed, then $X \prec_{l,c} CX$.

Proof

Take any $x \in X$, $y \notin X$. We have $y \preceq_c x$ or $x \prec_c y$, as any two elements are \preceq_c -comparable. $y \preceq_c x$ contradicts closure, so $x \prec_c y$, and $X \prec_{l,c} CX$ holds trivially. \square

2.4 Neighbourhood

Definition 2.9

Given a distance d , we define:

(1) Let $X \subseteq Y \subseteq U'$, then Y is a neighbourhood of X in U' iff

$$\forall y \in Y \forall x \in X (x \text{ is closest to } y \text{ among all } x' \text{ with } x' \in X \Rightarrow [x, y] \cap U' \subseteq Y).$$

(Closest means that there are no closer ones.)

When we also have a quality relation \prec , we define:

(2) Let $X \subseteq Y \subseteq U'$, then Y is an improving neighbourhood of X in U' iff

$$\forall y \in Y \forall x ((x \text{ is closest to } y \text{ among all } x' \text{ with } x' \in X \text{ and } x' \preceq y) \Rightarrow [x, y] \cap U' \subseteq Y).$$

When necessary, we will have to say for (3) and (4) which variant, i.e. set or counting, we mean.

Fact 2.10

(1) If $X \subseteq X' \subseteq \Sigma$, and $d(x, y) = 0 \Rightarrow x = y$, then X and X' are Hamming neighbourhoods of X in X' .

(2) If $X \subseteq Y_j \subseteq X' \subseteq \Sigma$ for $j \in J$, and all Y_j are Hamming Neighbourhoods of X in X' , then so are $\bigcup\{Y_j : j \in J\}$ and $\bigcap\{Y_j : j \in J\}$.

Proof

(1) is trivial (we need here that $d(x, y) = 0 \Rightarrow x = y$).

(2) Trivial.

\square

2.5 Unions of intersections and other definitions

Definition 2.10

Let \mathcal{O} over U be given.

$X \subseteq U'$ is (wi) (for union of intersections) iff there is a family $\mathcal{O}_i \subseteq \mathcal{O}$, $i \in I$ s.t. $X = (\bigcup\{\bigcap \mathcal{O}_i : i \in I\}) \cap U'$.

Unfortunately, this definition is not very useful for simple relativization.

Definition 2.11

Let \mathcal{O} be over U . Let $\mathcal{O}' \subseteq \mathcal{O}$. Define for $m \in U$ and $\delta : \mathcal{O}' \rightarrow 2 = \{0, 1\}$

$$m \models \delta \Leftrightarrow \forall O \in \mathcal{O}' (m \in O \Leftrightarrow \delta(O) = 1)$$

Definition 2.12

Let \mathcal{O} be over U .

\mathcal{O} is independent iff $\forall \delta : \mathcal{O} \rightarrow 2. \exists m \in U. m \models \delta$.

Obviously, independence does not inherit downward to subsets of U .

Definition 2.13

This definition is only intended for the set variant.

Let \mathcal{O} be over U .

$$\mathcal{D}(\mathcal{O}) := \{X \subseteq U' : \forall \mathcal{O}' \subseteq \mathcal{O} \forall \delta : \mathcal{O}' \rightarrow 2$$

$$((\exists m, m' \in U, m, m' \models \delta, m \in X, m' \notin X) \Rightarrow (\exists m'' \in X. m'' \models \delta \wedge m'' \prec_s m'))\}$$

This property expresses that we can satisfy obligations independently: If we respect \mathcal{O} , we can, in addition, respect \mathcal{O}' , and if we are hopeless kleptomaniacs, we may still not be a murderer. If $X \in \mathcal{D}(\mathcal{O})$, we can go from $U - X$ into X by improving on all $O \in \mathcal{O}$, which we have not fixed by δ , if δ is not too rigid.

3 Philosophical discussion of obligations

We take now a closer look at obligations, in particular at the ramifications of the treatment of the relation “better”. Some aspects of obligations will also need a notion of distance, we call them local properties of obligations.

3.1 A fundamental difference between facts and obligations: asymmetry and negation

There is an important difference between facts and obligations. A situation which satisfies an obligation is in some sense “good”, a situation which does not, is in some sense “bad”. This is not true of facts. Being “round” is a priori not better than “having corners” or vice versa. But given the obligation to post the letter, the letter in the mail box is “good”, the letter in the trash bin is “bad”. Consequently, negation has to play different role for obligations and for facts.

This is a fundamental property, which can also be found in orders, planning (we move towards the goal or not), reasoning with utility (is ϕ or $\neg\phi$ more useful?), and probably others, like perhaps the Black Raven paradox.

We also think that the Ross paradox (see below) is a true paradox, and should be avoided. A closer look shows that this paradox involves arbitrary weakening, in particular by the “negation” of an obligation. This was a starting point of our analysis.

“Good” and “bad” cannot mean that any situation satisfying obligation O is better than any situation not satisfying O , as the following example shows.

Example 3.1

If we have three independent and equally strong obligations, O , O' , O'' , then a situation satisfying O but neither O' nor O'' will not be better than one satisfying O' and O'' , but not O .

We have to introduce some kind of “ceteris paribus”. All other things being equal, a situation satisfying O is better than a situation not satisfying O , see Section 3.3 (page 11).

Example 3.2

The original version of the Ross paradox reads: If we have the obligation to post the letter, then we have the obligation to post or burn the letter. Implicit here is the background knowledge that burning the letter implies not to post it, and is even worse than not posting it.

We prefer a modified version, which works with two independent obligations: We have the obligation to post the letter, and we have the obligation to water the plants. We conclude by unrestricted weakening that we have the obligation to post the letter or *not* to water the plants. This is obvious nonsense.

It is not the “or” itself which is the problem. For instance, in case of an accident, to call an ambulance or to help the victims by giving first aid is a perfectly reasonable obligation. It is the negation of the obligation to water the plants which is the problem. More generally, it must not be that the system of suitable sets is closed under arbitrary supersets, otherwise we have closure under arbitrary right weakening, and thus the Ross paradox. Notions like “big subset” or “small exception sets” from the semantics of nonmonotonic logics are closed under supersets, so they are not suitable.

3.2 “And” and “or” for obligations

“Not” behaves differently for facts and for obligations. If O and O' are obligations, can $O \wedge O'$ be considered an obligation? We think, yes. “Ceteris paribus”, satisfying O and O' together is better than not to do so. If is the obligation to post the letter, O' to water the plants, then doing both is good, and better than doing none, or only one. Is $O \vee O'$ an obligation? Again, we think, yes. Satisfying one (or even both, a non-exclusive or) is better than doing nothing. We might not have enough time to do both, so we do our best, and water the plants or post the letter. Thus, if α and β are obligations, then so will be $\alpha \wedge \beta$ and $\alpha \vee \beta$, but not anything involving $\neg\alpha$ or $\neg\beta$. (In a non-trivial manner, leaving aside tautologies and contradictions which have to be considered separately.) To summarize: “and” and “or” preserve the asymmetry, “not” does not, therefore we can combine obligations using “and” and “or”, but not “not”. Thus, a reasonable notion of derivation of obligations will work with \wedge and \vee , but not with \neg .

We should not close under inverse \wedge , i.e. if $\phi \wedge \phi'$ is an obligation, we should not conclude that ϕ and ϕ' separately are obligations, as the following example shows.

Example 3.3

Let p stand for: post letter, w : water plants, s : strangle grandmother.

Consider now $\phi \wedge \phi'$, where $\phi = p \vee (\neg p \wedge \neg w)$, $\phi' = p \vee (\neg p \wedge w \wedge s)$. $\phi \wedge \phi'$ is equivalent to p - though it is perhaps a bizarre way to express the obligation to post the letter. ϕ leaves us the possibility not to water the plants, and ϕ' to strangle the grandmother, and neither seem good obligations. \square

Remark 3.1

This is particularly important in the case of soft obligations, as we see now, when we try to apply the rules of preferential reasoning to obligations.

One of the rules of preferential reasoning is the (OR) rule:

$$\phi \sim \psi, \phi' \sim \psi \Rightarrow \phi \vee \phi' \sim \psi.$$

Suppose we have $\phi \sim \psi' \wedge \psi''$, and $\phi' \sim \psi'$. We might be tempted to split $\psi' \wedge \psi''$ - as ψ' is a “legal” obligation, and argue: $\phi \sim \psi' \wedge \psi''$, so $\phi \sim \psi'$, moreover $\phi' \sim \psi'$, so $\phi \vee \phi' \sim \psi'$. The following example shows that this is not always justified.

Example 3.4

Consider the following obligations for a physician:

Let ϕ' imply that the patient has no heart disease, and if ϕ' holds, we should give drug A or (not drug A , but drug B), abbreviated $A \vee (\neg A \wedge B)$. (B is considered dangerous for people with heart problems.)

Let ϕ imply that the patient has heart problems. Here, the obligation is $(A \vee (\neg A \wedge B)) \wedge (A \vee (\neg A \wedge \neg B))$, equivalent to A .

The false conclusion would then be $\phi' \sim A \vee (\neg A \wedge B)$, and $\phi \sim A \vee (\neg A \wedge B)$, so $\phi \vee \phi' \sim A \vee (\neg A \wedge B)$, so in both situation we should either give A or B , but B is dangerous in “one half” of the situations.

□

We captured this idea about “and” and “or” in Definition 2.10 (page 9).

3.3 Ceteris paribus - a local property

Basically, the set of points “in” an obligation has to be better than the set of “exterior” points. As above Example 3.1 (page 10) with three obligations shows, demanding that any element inside is better than any element outside, is too strong. We use instead the “ceteris paribus” idea.

“All other things being equal” seems to play a crucial role in understanding obligations. Before we try to analyse it, we look for other concepts which have something to do with it.

The Stalnaker/Lewis semantics for counterfactual conditionals also works with some kind of “ceteris paribus”. “If it were to rain, I would use an umbrella” means something like: “If it were to rain, and there were not a very strong wind” (there is no such wind now), “if I had an umbrella” (I have one now), etc., i.e. if things were mostly as they are now, with the exception that now it does not rain, and in the situation I speak about it rains, then I will use an umbrella.

But also theory revision in the AGM sense contains - at least as objective - this idea: Change things as little as possible to incorporate some new information in a consistent way.

When looking at the “ceteris paribus” in obligations, a natural interpretation is to read it as “all other obligations being unchanged” (i.e. satisfied or not as before). This is then just a Hamming distance considering the obligations (but not other information).

Then, in particular, if \mathcal{O} is a family of obligations, and if x and x' are in the same subset $\mathcal{O}' \subseteq \mathcal{O}$ of obligations, then an obligation derived from \mathcal{O} should not separate them. More precisely, if $x \in \mathcal{O} \in \mathcal{O} \Leftrightarrow x' \in \mathcal{O} \in \mathcal{O}$, and D is a derived obligation, then $x \in D \Leftrightarrow x' \in D$.

Example 3.5

If the only obligation is not to kill, then it should not be derivable not to kill and to eat spaghetti.

Often, this is impossible, as obligations are not independent. In this case, but also in other situations, we can push “ceteris paribus” into an abstract distance d (as in the Stalnaker/Lewis semantics), which we postulate as given, and say that satisfying an obligation makes things better when going from “outside” the obligation to the d -closest situation “inside”. Conversely, whatever the analysis of “ceteris paribus”, and given a quality order on the situations, we can now define an obligation as a formula which (perhaps among other criteria) “ceteris paribus” improves the situation when we go from “outside” the formula “inside”.

A simpler way to capture “ceteris paribus” is to connect it directly to obligations, see Definition 2.13 (page 9). This is probably too much tied to independence (see below), and thus too rigid.

3.4 Hamming neighbourhoods

A combination concept is a Hamming neighbourhood:

X is called a Hamming neighbourhood of the best cases iff for any $x \in X$ and y a best case with minimal distance from x , all elements between x and y are in X .

For this, we need a notion of distance (also to define “between”). This was made precise in Definition 2.3 (page 5) and Definition 2.9 (page 9).

3.5 Global and mixed global/local properties of obligations

We look now at some global properties (or mixtures of global and local) which seem desirable for obligations:

(1) Downward closure

Consider the following example:

Example 3.6

Let $U' := \{x, x', y, y'\}$ with $x' := pqrs$, $y' := pqr\neg s$, $x := \neg p\neg qr\neg s$, $y := \neg p\neg q\neg r\neg s$.

Consider $X := \{x, x'\}$.

The counting version:

Then x' has quality 4 (the best), y' has quality 3, x has 1, y has 0.

$d_c(x', y') = 1$, $d_c(x, y) = 1$, $d_c(x, y') = 2$.

Then above “ceteris paribus” criterion is satisfied, as y' and x do not “see” each other, so $X \prec_{l,c} CX$.

But X is not downward closed, below $x \in X$ is a better element $y' \notin X$.

This seems an argument against X being an obligation.

The set version:

We still have $x' \prec_s y' \prec_s x \prec_s y$. As shown in Fact 2.6 (page 7), $d_s(x, y)$ (and also $d_s(x', y')$) and $d_s(x, y')$ are not comparable, so our argument collapses.

As a matter of fact, we have the result that the “ceteris paribus” criterion entails downward closure in the set variant, see Fact 2.8 (page 8).

□

Note that a sufficiently rich domain (put elements between y' and x) will make this local condition (for \prec) a global one, so we have here a domain problem. Domain problems are discussed e.g. in [Sch04] and [GS08a].

(2) Best states

It seems also reasonable to postulate that obligations contain all best states. In particular, obligations have then to be consistent - under the condition that best states exist. We are aware that this point can be debated, there is, of course, an easy technical way out: we take, when necessary, unions of obligations to cover the set of ideal cases. So obligations will be certain “neighbourhoods” of the “best” situations.

We think, that some such notion of neighbourhood is a good candidate for a semantics:

- A system of neighbourhoods is not necessarily closed under supersets.
- Obligations express something like an approximation to the ideal case where all obligations (if possible, or, as many as possible) are satisfied, so we try to be close to the ideal. If we satisfy an obligation, we are (relatively) close, and stay so as long as the obligation is satisfied.
- The notion of neighbourhood expresses the idea of being close, and containing everything which is sufficiently close. Behind “containing everything which is sufficiently close” is the idea of being in some sense convex. Thus, “convex” or “between” is another basic notion to be investigated. See here also the discussion of “between” in [Sch04].

3.6 Soft obligations

“Soft” obligations are obligations which have exceptions. Normally, one is obliged to do O , but there are cases where one is not obliged. This is like soft rules, as “Birds fly” (but penguins do not), where exceptions are not explicitly mentioned.

The semantic notions of size are very useful here, too. We will content ourselves that soft obligations satisfy the postulates of usual obligations everywhere except on a small set of cases. For instance, a soft obligation O should be downward closed “almost” everywhere, i.e. for a small subset of pairs $\langle a, b \rangle$ in $U \times U$ we accept that $a \prec b$, $b \in O$, $a \notin O$. We transplanted a suitable and cautious notion of size from the components to the product in Definition 2.1 (page 4).

When we look at the requirement to contain the best cases, we might have to soften this, too. We will admit that a small set of the ideal cases might be excluded. Small can be relative to all cases, or only to all ideal cases.

Soft obligations generate an ordering which takes care of exceptions, like the normality ordering of birds will take care of penguins: within the set of penguins, non-flying animals are the normal ones. Based on this ordering, we define “derived soft obligations”, they may have (a small set of) exceptions with respect to this ordering.

3.7 Overview of different types of obligations

(1) Hard obligations. They hold without exceptions, as in the Ten Commandments. You should not kill.

(1.1) In the simplest case, they apply everywhere and can be combined arbitrarily, i.e. for any $\mathcal{O}' \subseteq \mathcal{O}$ there is a model where all $O \in \mathcal{O}'$ hold, and no $O' \in \mathcal{O} - \mathcal{O}'$.

(1.2) In a more complicated case, not all combinations are possible. This is the same as considering just an arbitrary subset of U with the same set \mathcal{O} of obligations. This case is very similar to the case of conditional obligations (which might not be defined outside a subset of U), and we treat them together.

A good example is the Considerate Assassin:

Example 3.7

Normally, one should not offer a cigarette to someone, out of respect for his health. But the considerate assassin might do so nonetheless, on the cynical reasoning that the victim’s health is going to suffer anyway:

(1) One should not kill, $\neg k$.

(2) One should not offer cigarettes, $\neg o$.

(3) The assassin should offer his victim a cigarette before killing him, if k , then o .

Here, globally, $\neg k$ and $\neg o$ is best, but among k -worlds, o is better than $\neg o$. The model ranking is $\neg k \wedge \neg o \prec \neg k \wedge o \prec k \wedge o \prec k \wedge \neg o$.

Recall that an obligation for the whole set need not be an obligation for a subset any more, as it need not contain all best states. In this case, we may have to take a union with other obligations.

(2) Soft obligations.

Many obligations have exceptions. Consider the following example:

Example 3.8

You are in a library. Of course, you should not pour water on a book. But if the book has caught fire, you should pour water on it to prevent worse damage. In stenographic style these obligations read: “Do not pour water on books”. “If a book is on fire, do pour water on it.” It is like “birds fly”, but “penguins do not fly”, “soft” or nonmonotonic obligations, which have exceptions, which are not formulated in the original obligation, but added as exceptions.

We could have formulated the library obligation also without exceptions: “When you are in a library, and the book is not on fire, do not pour water on it.” “When you are in a library, and the book is on fire, pour water on it.” This formulation avoids exceptions. Conditional obligations behave like restricted quantifiers: they apply in a subset of all possible cases.

We treat now the considerate assassin case as an obligation (not to offer) with exceptions. Consider the full set U , and consider the obligation $\neg o$. This is not downward closed, as $k \wedge o$ is better than $k \wedge \neg o$. Downward closure will only hold for “most” cases, but not for all.

(3) Contrary-to-duty obligations.

Contrary-to-duty obligations are about different degrees of fulfillment. If you should ideally not have any fence, but are not willing or able to fulfill this obligation (e.g. you have a dog which might stray), then you should at least paint it white to make it less conspicuous. This is also a conditional obligation. Conditional, as it specifies what has to be done if there is a fence. The new aspect in contrary-to-duty obligations is the different degree of fulfillment.

We will not treat contrary-to-duty obligations here, as they do not seem to have any import on our basic ideas and solutions.

(4) A still more complicated case is when the language of obligations is not uniform, i.e. there are subsets $V \subseteq U$ where obligations are defined, which are not defined in $U - V$.

We will not pursue this case here.

3.8 Summary of the philosophical remarks

- (1) It seems justifiable to say that an obligation is satisfied or holds in a certain situation.
- (2) Obligations are fundamentally asymmetrical, thus negation has to be treated with care. “Or” and “and” behave as for facts.
- (3) Satisfying obligations improves the situation with respect to some given grading - *ceteris paribus*.
- (4) “*Ceteris paribus*” can be defined by minimal change with respect to other obligations, or by an abstract distance.
- (5) Conversely, given a grading and some distance, we can define an obligation locally as describing an improvement with respect to this grading when going from “outside” to the closest point “inside” the obligation.
- (6) Obligations should also have global properties: they should be downward (i.e. under increasing quality) closed, and cover the set of ideal cases.
- (7) The properties of “soft” obligations, i.e. with exceptions, have to be modified appropriately. Soft obligations generate an ordering, which in turn may generate other obligations, where exceptions to the ordering are permitted.
- (8) Quality and distance can be defined from an existing set of obligations in the set or the counting variant. Their behaviour is quite different.
- (9) We distinguished various cases of obligations, soft and hard, with and without all possibilities, etc.

Finally, we should emphasize that the notions of distance, quality, and size are in principle independent, even if they may be based on a common substructure.

4 Examination of the various cases

We will concentrate here on the set version of hard obligations.

4.1 Hard obligations for the set approach

4.1.1 Introduction

We work here in the set version, the \in -case, and examine mostly the set version only.

We will assume a set \mathcal{O} of obligations to be given. We define the relation $\prec := \prec_{\mathcal{O}}$ as described in Definition 2.3 (page 5), and the distance d is the Hamming distance based on \mathcal{O} .

4.1.2 The not necessarily independent case

Example 4.1

Work in the set variant. We show that $X \preceq_s$ -closed does not necessarily imply that X contains all \preceq_s -best elements. Let $\mathcal{O} := \{p, q\}$, $U' := \{p \neg q, \neg pq\}$, then all elements of U' have best quality in U' , $X := \{p \neg q\}$ is closed, but does not contain all best elements. \square

Example 4.2

Work in the set variant. We show that $X \preceq_s$ -closed does not necessarily imply that X is a neighbourhood of the best elements, even if X contains them.

Consider $x := pq \neg rstu$, $x' := \neg pqrs \neg t \neg u$, $x'' := p \neg qr \neg s \neg t \neg u$, $y := p \neg q \neg r \neg s \neg t \neg u$, $z := pq \neg r \neg s \neg t \neg u$. $U := \{x, x', x'', y, z\}$, the \prec_s -best elements are x, x', x'' , they are contained in $X := \{x, x', x'', z\}$. $d_s(z, x) = \{s, t, u\}$, $d_s(z, x') = \{p, r, s\}$, $d_s(z, x'') = \{q, r\}$, so x'' is one of the best elements closest to z . $d(z, y) = \{q\}$, $d(y, x'') = \{r\}$, so $[z, x''] = \{z, y, x''\}$, $y \notin X$, but X is downward closed. \square

Fact 4.1

Work in the set variant.

Let $X \neq \emptyset$, $X \preceq_s$ -closed. Then

- (1) X does not necessarily contain all best elements.
- Assume now that X contains, in addition, all best elements. Then
- (2) $X \prec_{l,s} CX$ does not necessarily hold.
- (3) X is (ui).
- (4) $X \in \mathcal{D}(\mathcal{O})$ does not necessarily hold.
- (5) X is not necessarily a neighbourhood of the best elements.
- (6) X is an improving neighbourhood of the best elements.

Proof

- (1) See Example 4.1 (page 14)
- (2) See Example 2.2 (page 8)
- (3) If there is $m \in X$, $m \notin O$ for all $O \in \mathcal{O}$, then by closure $X = U$, take $\mathcal{O}_i := \emptyset$.

For $m \in X$ let $\mathcal{O}_m := \{O \in \mathcal{O} : m \in O\}$. Let $X' := \bigcup \{\bigcap \mathcal{O}_m : m \in X\}$.

$X \subseteq X'$: trivial, as $m \in X \rightarrow m \in \bigcap \mathcal{O}_m \subseteq X'$.

$X' \subseteq X$: Let $m' \in \bigcap \mathcal{O}_m$ for some $m \in X$. It suffices to show that $m' \preceq_s m$. $m' \in \bigcap \mathcal{O}_m = \bigcap \{O \in \mathcal{O} : m \in O\}$, so for all $O \in \mathcal{O}$ ($m \in O \rightarrow m' \in O$).

(4) Consider Example 2.2 (page 8), let $\text{dom}(\delta) = \{r, s\}$, $\delta(r) = \delta(s) = 0$. Then $x, y \models \delta$, but $x' \not\models \delta$ and $x \in X$, $y \notin X$, but there is no $z \in X$, $z \models \delta$ and $z \prec y$, so $X \notin \mathcal{D}(\mathcal{O})$.

- (5) See Example 4.2 (page 14).
- (6) By Fact 2.6 (page 7), (5).

\square

Fact 4.2

Work in the set variant

- (1.1) $X \prec_{l,s} CX$ implies that X is \preceq_s -closed.
- (1.2) $X \prec_{l,s} CX \Rightarrow X$ contains all best elements
- (2.1) X is (ui) $\Rightarrow X$ is \preceq_s -closed.
- (2.2) X is (ui) does not necessarily imply that X contains all \preceq_s -best elements.
- (3.1) $X \in \mathcal{D}(\mathcal{O}) \Rightarrow X$ is \preceq_s -closed
- (3.2) $X \in \mathcal{D}(\mathcal{O})$ implies that X contains all \preceq_s -best elements.
- (4.1) X is an improving neighbourhood of the \preceq_s -best elements $\Rightarrow X$ is \preceq_s -closed.
- (4.2) X is an improving neighbourhood of the best elements $\Rightarrow X$ contains all best elements.

Proof

- (1.1) By Fact 2.8 (page 8).

(1.2) By Fact 2.7 (page 8).

(2.1) Let $O \in \mathcal{O}$, then O is downward closed (no $y \notin O$ can be better than $x \in O$). The rest follows from Fact 2.5 (page 6) (3).

(2.2) Consider Example 4.1 (page 14), p is (ui) (formed in $U!$), but $p \cap X$ does not contain $\neg pq$.

(3.1) Let $X \in \mathcal{D}(\mathcal{O})$, but let X not be closed. Thus, there are $m \in X$, $m' \preceq_s m$, $m' \notin X$.

Case 1: Suppose $m' \sim m$. Let $\delta_m : \mathcal{O} \rightarrow 2$, $\delta_m(O) = 1$ iff $m \in O$. Then $m, m' \models \delta_m$, and there cannot be any $m'' \models \delta_m$, $m'' \prec_s m'$, so $X \notin \mathcal{D}(\mathcal{O})$.

Case 2: $m' \prec_s m$. Let $\mathcal{O}' := \{O \in \mathcal{O} : m \in O \Leftrightarrow m' \in O\}$, $\text{dom}(\delta) = \mathcal{O}'$, $\delta(O) := 1$ iff $m \in O$ for $O \in \mathcal{O}'$. Then $m, m' \models \delta$. If there is $O \in \mathcal{O}$ s.t. $m' \notin O$, then by $m' \preceq_s m$ $m \notin O$, so $O \in \mathcal{O}'$. Thus for all $O \notin \text{dom}(\delta)$, $m' \in O$. But then there is no $m'' \models \delta$, $m'' \prec_s m'$, as m' is already optimal among the n with $n \models \delta$.

(3.2) Suppose $X \in \mathcal{D}(\mathcal{O})$, $x' \in U - X$ is a best element, take $\delta := \emptyset$, $x \in X$. Then there must be $x'' \prec x'$, $x'' \in X$, but this is impossible as x' was best.

(4.1) By Fact 2.6 (page 7), (4) all minimal elements have incomparable distance. But if $z \preceq y$, $y \in X$, then either z is minimal or it is above a minimal element, with minimal distance from y , so $z \in X$ by Fact 2.6 (page 7) (3).

(4.2) Trivial.

□

4.1.3 The independent case

Assume now the system to be independent, i.e. all combinations of \mathcal{O} are present.

Note that there is now only one minimal element, and the notions of Hamming neighbourhood of the best elements and improving Hamming neighbourhood of the best elements coincide.

Fact 4.3

Work in the set variant.

Let $X \neq \emptyset$, $X \preceq_s$ -closed. Then

- (1) X contains the best element.
- (2) $X \prec_{l,s} CX$
- (3) X is (ui).
- (4) $X \in \mathcal{D}(\mathcal{O})$
- (5) X is a (improving) Hamming neighbourhood of the best elements.

Proof

(1) Trivial.

(2) Fix $x \in X$, let y be closest to x , $y \notin X$. Suppose $x \not\prec y$, then there must be $O \in \mathcal{O}$ s.t. $y \in O$, $x \notin O$. Choose y' s.t. y' is like y , only $y' \notin O$. If $y' \in X$, then by closure $y \in X$, so $y' \notin X$. But y' is closer to x than y is, *contradiction*.

Fix $y \in U - X$. Let x be closest to y , $x \in X$. Suppose $x \not\prec y$, then there is $O \in \mathcal{O}$ s.t. $y \in O$, $x \notin O$. Choose x' s.t. x' is like x , only $x' \in O$. By closure of X , $x' \in X$, but x' is closer to y than x is, *contradiction*.

(3) By Fact 4.1 (page 14) (3)

(4) Let X be closed, and $\mathcal{O}' \subseteq \mathcal{O}$, $\delta : \mathcal{O}' \rightarrow 2$, $m, m' \models \delta$, $m \in X$, $m' \notin X$. Let m'' be s.t. $m'' \models \delta$, and for all $O \in \mathcal{O} - \text{dom}(\delta)$ $m'' \in O$. This exists by independence. Then $m'' \preceq_s m'$, but also $m'' \preceq_s m$, so $m'' \in X$. Suppose $m'' \sim m'$, then $m' \preceq_s m''$, so $m' \in X$, contradiction, so $m'' \prec_s m'$.

(5) Trivial by (1), the remark preceding this Fact, and Fact 4.1 (page 14) (6).

Fact 4.4

Work in the set variant.

- (1) $X \prec_{l,s} Cx \Rightarrow X$ is \preceq_s -closed,
- (2) X is (ui) $\Rightarrow X$ is \preceq_s -closed,
- (3) $X \in \mathcal{D}(\mathcal{O}) \Rightarrow X$ is \preceq_s -closed,
- (4) X is a (improving) neighbourhood of the best elements $\Rightarrow X$ is \preceq_s -closed.

Proof

(1) Suppose there are $x \in X$, $y \in U - X$, $y \prec x$. Choose them with minimal distance. If $\text{card}(d_s(x, y)) > 1$, then there is z , $y \prec_s z \prec_s x$, $z \in X$ or $z \in U - X$, contradicting minimality. So $\text{card}(d_s(x, y)) = 1$. So y is among the closest elements of $U - X$ seen from x , but then by prerequisite $x \prec y$, *contradiction*.

(2) By Fact 4.2 (page 14) (2.1).

(3) By Fact 4.2 (page 14) (3.1).

(4) There is just one best element z , so if $x \in X$, then $[x, z]$ contains all $y \prec x$ by Fact 2.6 (page 7) (3).

□

The $\mathcal{D}(\mathcal{O})$ condition seems to be adequate only for the independent situation, so we stop considering it now.

Fact 4.5

Let $X_i \subseteq U$, $i \in I$ a family of sets, we note the following about closure under unions and intersections:

- (1) If the X_i are downward closed, then so are their unions and intersections.
- (2) If the X_i are (ui), then so are their unions and intersections.

Proof

Trivial. □

We do not know whether $\prec_{l,s}$ is preserved under unions and intersections, it does not seem an easy problem.

Fact 4.6

- (1) Being downward closed is preserved while going to subsets.
- (2) Containing the best elements is not preserved (and thus neither the neighbourhood property).
- (3) The $\mathcal{D}(\mathcal{O})$ property is not preserved.
- (4) $\preceq_{l,s}$ is not preserved.

Proof

(4) Consider Example 3.6 (page 11), and eliminate y from U' , then the closest to x not in X is y' , which is better.

□

4.2 Remarks on the counting case

Remark 4.7

In the counting variant all qualities are comparable. So if X is closed, it will contain all minimal elements.

Example 4.3

We measure distance by counting.

Consider $a := \neg p \neg q \neg r \neg s$, $b := \neg p \neg q \neg r s$, $c := \neg p \neg q r \neg s$, $d := p q r \neg s$, let $U := \{a, b, c, d\}$, $X := \{a, c, d\}$. d is the best element, $[a, d] = \{a, d, c\}$, so X is an improving Hamming neighbourhood, but $b \prec a$, so $X \not\prec_{l,c} CX$.

□

Fact 4.8

We measure distances by counting.

$X \prec_{l,c} CX$ does not necessarily imply that X is an improving Hamming neighbourhood of the best elements.

Proof

Consider Example 3.6 (page 11). There $X \prec_{l,c} CX$. x' is the best element, and $y' \in [x', x]$, but $y' \notin X$. □

5 What is an obligation?

The reader will probably not expect a final definition. All we can do is to give a tentative definition, which, in all probability, will not be satisfactory in all cases.

Definition 5.1

We decide for the set relation and distance.

- (1) Hard obligation

A hard obligation has the following properties:

- (1.1) It contains all ideal cases in the set considered.

(1.2) It is closed under increasing quality, Definition 2.6 (page 6)

(1.3) It is an improving neighbourhood of the ideal cases (this also implies (1.1)), Definition 2.9 (page 9)

We are less committed to:

(1.4) It is *ceteris paribus* improving, Definition 2.8 (page 8)

An obligation O is a derived obligation of a system \mathcal{O} of obligations iff it is a hard obligation based on the set variant of the order and distance generated by \mathcal{O} .

(2) Soft obligations

A set is a soft obligation iff it satisfies the soft versions of above postulates. The notion of size has to be given, and is transferred to products as described in Definition 2.1 (page 4). More precisely, strict universal quantifiers are transformed into their soft variant “almost all”, and the other operators are left as they are. Of course, one might also want to use a mixture of soft and hard conditions, e.g. we might want to have all ideal cases, but renounce on closure for a small set of pairs $\langle x, x' \rangle$.

An obligation O is derived from \mathcal{O} iff it is a soft obligation based on the set variant of the order and distance generated by the translation of \mathcal{O} into their hard versions. (I.e. exceptions will be made explicit.)

Fact 5.1

Let $O \in \mathcal{O}$, then $\mathcal{O} \sim O$ in the independent set case.

Proof

We check (1.1) – (1.3) of Definition 5.1 (page 16).

(1.1) holds by independence.

(1.2) If $x \in O$, $x' \notin O$, then $x' \not\prec_s x$.

(1.3) By Fact 4.1 (page 14) (6).

Note that (1.4) will also hold by Fact 4.3 (page 15) (2).

□

Corollary 5.2

Every derived obligation is a classical consequence of the original set of obligations in the independent set case.

Proof

This follows from Fact 4.3 (page 15) (3) and Fact 5.1 (page 17).

Example 5.1

The Ross paradox is not a derived obligation.

Proof

Suppose we have the alphabet p, q and the obligations $\{p, q\}$, let $R := p \vee \neg q$. This is not not closed, as $\neg p \wedge q \prec \neg p \wedge \neg q \in R$. □

6 Conclusion

Obligations differ from facts in the behaviour of negation, but not of conjunction and disjunction. The Ross paradox originates, in our opinion, from the differences in negation. Central to the treatment of obligations seems to be a relation of “better”, which can generate obligations, but also be generated by obligations. The connection between obligations and this relation of “better” seems to be somewhat complicated and leads to a number of ramifications. A tentative definition of a derivation of obligations is given.

7 Acknowledgements

We thank A.Herzig, Toulouse, and L.v.d.Torre, Luxembourg, for very helpful discussions.

References

- [CJ02] J.Carmo, A.J.I.Jones, “Deontic logic and contrary-to-duties”, in: Handbook of Philosophical Logic, Vol. 8, D.Gabbay, F.Guenther eds., pp. 265-343, Kluwer, 2002
- [GS08a] D.Gabbay, K.Schlechta, “Cumulativity without closure of the domain under finite unions”, hal-00311938, arXiv 0808.3077, to appear in Review of Symbolic Logic

- [GS09a] D.Gabbay, K.Schlechta, “Size and logic”, submitted, see also arXiv 0903.1367
- [Han69] B.Hansson, “An analysis of some deontic logics”, *Nous* 3, 373-398. Reprinted in R.Hilpinen ed. “Deontic Logic: Introductory and Systematic Readings”. Reidel, Dordrecht 1971, 121-147
- [MDW94] J.J.Ch.Meyer, F.P.M.Dignum, R.J.Wieringa, “The paradoxes of deontic logic revisited: a computer science perspective”, University of Utrecht, NL, Dept. Comp.Sc., Tech.Rep., 1994
- [Mon70] R.Montague, “Universal grammar”, *Theoria* 36, 373-98, 1970,
- [Pac07] O.Pacheco, “Neighbourhood semantics for deontic and agency logics” *CIC'07*, October 2007
- [Sch04] K.Schlechta: “Coherent Systems”, Elsevier, Amsterdam, 2004
- [Sco70] D.Scott: “Advice in modal logic”, in: “Philosophical problems in Logic”, K.Lambert ed., Reidel 1970