# Learning and Diachronic Laws for Partial Blocking

**Anton Benz**
**Syddansk Universitet, Kolding**

**Abstract**

In this talk we introduce a complete system of diachronic laws for predicting partial blocking. The laws get their justification from an underlying learning model, and the system is *complete* in the sense that all possible meaning shifts predicted by the learning model can be accounted for by using diachronic laws instead. We propose them as an alternative to Horn's division of pragmatic labour and the principle of weak optimality.
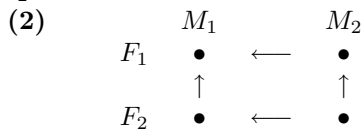
## 1 Introduction

Horn's principle of division of pragmatic labour [3] states that marked forms have a tendency to go together with marked meanings, and unmarked forms with unmarked meanings. This accounts for partial blocking phenomena as observed in the following examples:

**(1) a)** John mopped the floor with water / a liquid.
    **b)** Black Bart killed the sheriff / caused the sheriff to die.
    **c)** Two Americans / Two Latin-Americans have been killed in the plot.

Normally, people use water for mopping a floor, hence the use of the marked form *liquid* indicates that it is not *water* what John used. Normal killing–events are events of direct killing, hence the use of the marked form *cause to die* indicates that it was not a direct killing. The use of the unmarked form *Americans* indicates that US–Americans have been killed.

In general, if $F_1$ and $F_2$ are forms and $M_1$ and $M_2$ are meanings where $F_1$ is preferred over $F_2$ and $M_1$ over $M_2$, then $F_1$ tends to denote $M_1$ and $F_2$ to denote $M_2$:

**(2)**

$$
\begin{array}{ccc}
 & M_1 & M_2 \\
F_1 & \bullet \longleftarrow & \bullet \\
 & \uparrow & \uparrow \\
F_2 & \bullet \longleftarrow & \bullet
\end{array}
$$

Graphs like **(2)** are familiar from Bidirectional Optimality Theory. Blutner's principle of *weak optimality*, or *superoptimality*, is a reformulation and generalisation of Horn's principle in an optimality theoretic framework [2]. A drawback of both principles is their tendency to over–generate blocking phenomena. E.g. for **(1)** b) they predict not only partial blocking for *kill* and *cause to die*, but also for *cause to die* and *made to be killed* and any other more complex phrase classifying a killing–event.

In this talk we are going to explain partial blocking by diachronic laws derived from an underlying learning model. The fundamental learning principle of

this model can be summarised as follows: If in every situation where a form $F$ with meaning $\mathbf{f}$ is used for classifying some entity it turns out that this entity is of a stronger type $\mathbf{t} \leq \mathbf{f}$, then the language users will learn to use $F$ as meaning $\mathbf{t}$. This strengthened meaning remains defeasible in principle, hence we call it *associated* meaning.

What do we mean by *diachronic law*? Let $\{F_0, \ldots, F_n\}$ be a set of forms. We assume that they can be linearly ordered according to their complexity. Then a diachronic law will have the form: If in a diachronic stage $i$ the semantic relations between $F_0, \ldots, F_n$ are such and such and there occur only entities of type $\mathbf{t}_0, \ldots, \mathbf{t}_m$, then the semantic relations in stage $i + 1$ will be these and that.

As an example we consider (**1**) a). We can simplify and assume that there are only entities of two types: $\mathbf{t}_0 = +water$ and $\mathbf{t}_1 = -water$. Further we can assume that there are only three forms to be considered: $F_0 = water$, $F_1 = something\ that\ is\ not\ water$, and $F_2 = liquid$. $F_0$ is less marked than $F_2$, and $F_2$ less marked than $F_1$. When is there a reason to use $F_2 = liquid$ for classifying something that John uses for mopping the floor? If it is water, then the speaker will see that it is water, and hence the choice of the form *water* is most economic. There will never be a reason to use *liquid*, only if, in fact, it is something different from water what John uses. The above learning rule implies that the form *liquid* gets associated with the meaning $-water$. This can be turned into a law:

(**A**) If in stage $i$ $F_0$ is the most economic form with meaning $\mathbf{t}_0$, $F_1$ with meaning $\mathbf{t}_1$, and $F_2$ with meaning $\mathbf{t}_0 \vee \mathbf{t}_1$, and if $F_0 < F_2 < F_1$, then in stage $i + 1$ $F_2$ is associated with $\mathbf{t}_1$.

This *law* does not make reference to the type of entities occurring in stage $i$. In (**1**) b) the reason why *kill* gets associated with *direct killing* seems to be that normally only direct killings occur. The fact that only some types are realised gives rise to another list of laws. We will present a complete list. Fortunately, there is only a small number: If we concentrate on the case for two basic types $\mathbf{t}_0, \mathbf{t}_1$ as in (**1**) a), then there are in addition to (**A**) only five laws describing all possible ways of how strengthening of meaning can develop.

## 2 Diachronic Laws in the Situation with two Basic Types

We promised to explain partial blocking by diachronic laws derived from an underlying learning model. First, we provide for a classification of utterance situations where the speaker has to make a choice between forms. Then we shortly introduce the formal learning model. Finally, we show how to derive diachronic laws from this model and use them for determining how and when Horn situations can develop out of Blutner–squares (**2**).

Is there a complete characterisation of all possible diachronic processes in terms of laws of diachronic change? Given a set of semantically synonymous expressions, how and when can associative learning and speaker's preferences lead to a change in interpretation? We work out an answer for the situation with two basic types.

### 2.1 The Classification of Utterance Situations

We make the following assumptions about the utterance situations: The speaker wants to classify some object or event $e$ as being of some type $\mathbf{f}$. It is common ground that he knows $e$. Hence we represent an utterance situation where the speaker has to

make his choice for a form $F$ by a pair $\langle e, \mathbf{f} \rangle$. Lets assume that the classified entities can differ only with respect to one feature and that attribute–value functions that represent the meanings can have only three values, namely $\{-1, 0, 1\}$. Let $m$ be a feature and $[i] := \{F \in \mathcal{F} \mid [F](m) = i\}$, where $[F]$ is the attribute–value function representing the meaning of $F$. Let $\preceq$ be a linear well–founded order on $\mathcal{F}$ with meaning: $F \prec F'$ iff $F'$ is more complex than $F$. It follows that for each $[i]$ there is a unique minimal form in $[i]$. As the speaker will choose the most preferred form, he has to consider only three forms: The minimal elements of $[-1]$, $[0]$ and $[1]$.

In general, if we consider a situation with two basic types $\mathbf{t}_0$ and $\mathbf{t}_1$, then there are only three forms $F_0, F_1, F_2$ the speaker has to consider for making his choice. Without loss of generality we can assume that $[F_0] = \mathbf{t}_0$, $[F_1] = \mathbf{t}_1$ and $[F_2] = \mathbf{t}_0 \vee \mathbf{t}_1$. Hence, $F_2$ always denotes the form with the wider meaning. We can further assume that in general $F_0$ is preferred over $F_1$. Hence, we arrive at the following complete classification of all choice situations with two basic types:



Case I       Case II       Case III
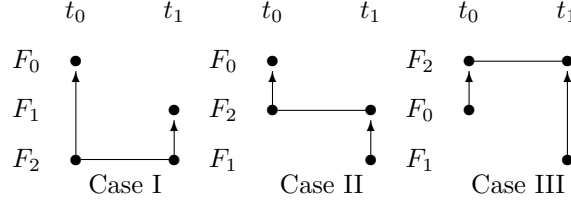
The topmost form is the most preferred one, the lowest the least preferred. The vertical arrows indicate the speaker's preferences. The horizontal line means that the respective form has an extension which comprises the meaning of both types $\mathbf{t}_0$ and $\mathbf{t}_1$. Examples are: Case I *father, mother, one of the parents* $(F_0 \prec F_1 \prec F_2)$; Case II *water, liquid, alcoholic essence* $(F_0 \prec F_2 \prec F_1)$; Case III American, North American, Latin American $(F_2 \prec F_0 \prec F_1)$. If $F_0$ and $F_1$ are adjacent, then the relation between their complexities is irrelevant. Future classifications of concrete examples is meant up to renaming of types and forms.

## 2.2 Associative Learning

We represent a *diachronic stage* by a triple $\langle E, S, H \rangle$: $E$ is a set of utterance situations of the form $\langle e, \mathbf{f} \rangle$; $S$ is a function from $E$ into forms $\mathcal{F}$ and represents the speaker's choice in all situations in $E$; and $H$ is a function from $\mathcal{F}$ into types and represents the hearer's interpretation of forms. The speaker's choice $S(e, \mathbf{f})$ of a form $F$ is *successful* in $\langle e, \mathbf{f} \rangle$ if $e : H(F) \& H(F) \leq \mathbf{f}$, i.e. if $e$ is of type $H(F) = H(S(e, \mathbf{f}))$ and if the hearer can therefore infer that it is of type $\mathbf{f}$.

We present a model for associative learning as described above. Assume we are in stage $\langle E, S, H \rangle$. How does the new selection and interpretation strategies in the next stage look like? The basic ideas are:

- The hearer learns that the factual information of an utterance is stronger than its semantics.
- The speaker learns to exploit this situation.

Let us assume that $F$ is a form that the speaker uses in the given stage; then:

$$H^+(F) \quad := \quad \min\{\mathbf{f} \in \mathbf{Type} \mid \mathbf{f} \leq H(F) \wedge \|F\| \subseteq [\![\mathbf{f}]\!]\} \tag{2.1}$$

$$S^+(e, \mathbf{f}) \quad := \quad \min\{F \in NL \mid e : H^+(F) \leq \mathbf{f}\}. \tag{2.2}$$

$[\![\mathbf{f}]\!]$ denotes the *extension* of $\mathbf{f}$ in $E$, i.e. $[\![\mathbf{f}]\!] := \{e \in E \mid e : \mathbf{f}\}$. $\|F\|$ is the set of all entities where the speaker has in fact used $F$ to classify them, i.e.

$$\|F\| := \{e \in E \mid \exists \mathbf{f} : \langle e, \mathbf{f} \rangle \in E \wedge S(e, \mathbf{f}) = F\}. \tag{2.3}$$
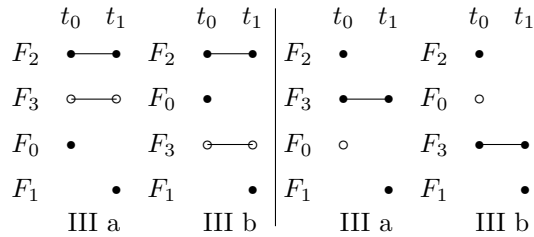
For **(1)** a) this means: As there is only a reason to use *liquid* if the classified entity is not water, it follows that $\|liquid\| \subseteq \{e \in E \mid e : -water\} = [\![-water]\!]$. As there is no stronger type than $-water$, we find $H^+(liquid) = -water$. For **(1)** b) we find: In the initial stage only direct killings occur; it follows that $\|kill\| \subseteq \{e \in E \mid e : direct\ killing\}$. Hence, $H^+(kill) = directly\ killing$. In the following stage there will be a reason to use the form $F' = cause\ to\ die$ only if an hitherto unusual *indirect* killing event occurs. It follows that $H^{++}(F') = indirectly\ killing$.

   If $F$ is not used in the given stage, then no strengthening should occur; i.e. $H^+(F) := H(F)$. $H^+$ and $S^+$ describe both, the hearer's and the speaker's learning[1]. The hearer's learning precedes the speaker's, but we put both processes together in one stage[2].

   As long as we consider only isolated examples, the associative learning model may be sufficient for explaining the observed data. But if we ask for overall regularities, then it is a great advantage to start with a classification of (1) dialogue situations, as done in Sec. 2.1, and (2) laws that describe how these situations can develop diachronically.

## 2.3   Laws of Diachronic Change

We restrict our considerations further to situations where there is for each type $\mathbf{t}_i$ a situation where the speaker wants to classify the object only as $\mathbf{t}_0 \vee \mathbf{t}_1$; i.e. if $e \in [\![\mathbf{t}_i]\!]$, then $\langle e, \mathbf{t}_0 \vee \mathbf{t}_1 \rangle \in E$. What parameters can change diachronically? Beside selection and interpretation strategies, there is only one: The set $E$ of utterance situations. As there are only two basic types, $\mathbf{t}_0$ and $\mathbf{t}_1$, there are only two possibilities how reduced occurrences of entities can have an influence within the associative learning model: Either type $\mathbf{t}_0$ or $\mathbf{t}_1$ is not realised in $E$. If only $\mathbf{t}_0$ is realised, we say that we get the new situation by $\{\mathbf{t}_0\}$–reduction, and if only $\mathbf{t}_1$ is realised, we say that we get the new situation by $\{\mathbf{t}_1\}$–reduction. It is possible that a $\{\mathbf{t}_1\}$–reduction follows a $\{\mathbf{t}_0\}$–reduction: We see reduction always relative to the full situation given by Case I to Case III. $\{\mathbf{t}_i\}$–reduction has the effect that the hearer associates $\mathbf{t}_i$ with the lightest form $F_j$ that could classify $\mathbf{t}_i$–entities. Lets consider the situation for Case III examples. Let $F_3$ be another form with wide meaning but more complex than $F_2$. Which effects has $\{\mathbf{t}_0\}$–reduction? There are only three possible types of situations: Either (a) $F_2 \prec F_3 \prec F_0 \prec F_1$, (b) $F_2 \prec F_0 \prec F_3 \prec F_1$, or (c) $F_2 \prec F_0 \prec F_1 \prec F_3$. For (a) and (b) the situation looks as follows (left side):

| | $t_0$ | $t_1$ | | $t_0$ | $t_1$ | | $t_0$ | $t_1$ | | $t_0$ | $t_1$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $F_2$ | ●—— | ● | $F_2$ | ●—— | ● | $F_2$ | ● | | $F_2$ | ● | |
| $F_3$ | ○—— | ○ | $F_0$ | ● | | $F_3$ | ●—— | ● | $F_0$ | ○ | |
| $F_0$ | ● | | $F_3$ | ○—— | ○ | $F_0$ | ○ | | $F_3$ | ●—— | ● |
| $F_1$ | | ● | $F_1$ | | ● | $F_1$ | | ● | $F_1$ | | ● |
| | III a | | | III b | | | III a | | | III b | |

The hollow bullets mean that the speaker has never a reason to choose the respective form. $\{\mathbf{t}_0\}$–reduction means that the hearer learns to associate $\mathbf{t}_0$ with the least complex form $F_2$. The situation resulting from learning is depicted at the right side. We see that a Case II situation has emerged. For (c) $\{\mathbf{t}_0\}$–reduction would

---

1. The learning model is related to *classifier learning* [5].
2. For more information on the associative learning model see [1].

lead to a Case I situation. **(1)** b) is an example for III (a), and if we set $F_3 :=$ *Inhabitant of the American continent*, then **(1)** c) is an example for III (c).

For Case II there can only be two further sub–cases: Either (a) $F_0 \prec F_2 \prec F_3 \prec F_1$, or (b) $F_0 \prec F_2 \prec F_1 \prec F_3$. For Case I there is only one: $F_0 \prec F_1 \prec F_2 \prec F_3$. Reduction and subsequent associative learning yields the following list of laws:
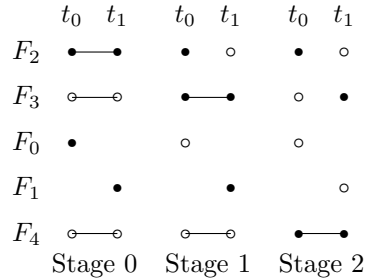
*Reduction Laws:*

**(R1)** II situations turn by $\{\mathbf{t}_1\}$–reduction into I situations where $F_2$ is associated with $\mathbf{t}_1$ and $F_3$ is the lightest expression with meaning $\mathbf{t}_0 \vee \mathbf{t}_1$.

**(R2)** III a) situations turn by $\{\mathbf{t}_i\}$–reduction into II situations where $F_2$ is associated with $\mathbf{t}_i$ and $F_3$ is the lightest expression with meaning $\mathbf{t}_0 \vee \mathbf{t}_1$.

**(R3)** III b) situations turn by $\{\mathbf{t}_0\}$–reduction into II situations where $F_2$ is associated with $\mathbf{t}_0$ and $F_3$ is the lightest expression with meaning $\mathbf{t}_0 \vee \mathbf{t}_1$.

**(R4)** III b) situations turn by $\{\mathbf{t}_1\}$–reduction into I situations where $F_2$ is associated with $\mathbf{t}_1$ and $F_3$ is the lightest expression with meaning $\mathbf{t}_0 \vee \mathbf{t}_1$.

**(R5)** III c) situations turn by $\{\mathbf{t}_i\}$–reduction into I situations where $F_2$ is associated with $\mathbf{t}_i$ and $F_3$ is the lightest expression with meaning $\mathbf{t}_0 \vee \mathbf{t}_1$.

The classification of the resulting state is again meant to be correct up to suitable renaming. The effect of $\{\mathbf{t}_1\}$–reduction in Case II situations is the same as the effect of simple associative learning without reduction. Hence, **(R1)** is covered by the following law:

*Law of Associative Learning:*

**(A)** Case II situations turn into Case I situations where $F_2$ is associated with $\mathbf{t}_1$ and $F_3$ is the lightest expression with meaning $\mathbf{t}_0 \vee \mathbf{t}_1$.
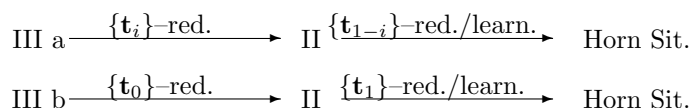
In all other cases the resulting situation is the same as the original one. Now it is not difficult to see how and when we can derive the effects of Horn's division of pragmatic labour for Blutner-squares **(2)**. We need an initial situation with two co–extensive forms $F_2$ and $F_3$ which can develop into a Case I situation where $F_2$ is interpreted either as $\mathbf{t}_0$ or $\mathbf{t}_1$, and $F_3$ as the other one. There are only two such situations: Case III a) and Case III b) situations. For Case III a) the desired Case I situation emerges by a three–stage process:



The second reduction law **(R2)** implies that the situation on the left side turns into the situation in the middle by $\{\mathbf{t}_0\}$–reduction. The case for $\{\mathbf{t}_1\}$–reduction is symmetric. Then, either by the first reduction law, or by the law of associative learning, the situation in the middle turns into the situation on the right. The hollow bullets in the rows for $F_2$ and $F_3$ should indicate that the respective type is still part of the semantic meaning of the form but it is excluded from its actual default interpretation. The case for III b) differs from III a) because the first reduction

must be a $\{\mathbf{t}_0\}$–reduction. The first two rows of Stage 0 form a Blutner square **(2)**. Let us call a situation like that represented by the first two rows in Stage 2 a *Horn situation*. Blutner's principle of weak optimality diachronically interpreted predicts that the Blutner square in Stage 0 develops into a Horn situation. We can recover this principle as follows:

*The Emergence of Horn Situations:   A Horn situation can only develop out of III a) and III b) examples. It emerges as the result of the following two processes:*

$$\text{III a} \xrightarrow{\{\mathbf{t}_i\}\text{–red.}} \text{II} \xrightarrow{\{\mathbf{t}_{1-i}\}\text{–red./learn.}} \text{Horn Sit.}$$

$$\text{III b} \xrightarrow{\{\mathbf{t}_0\}\text{–red.}} \text{II} \xrightarrow{\{\mathbf{t}_1\}\text{–red./learn.}} \text{Horn Sit.}$$

We can also see the result of turning a Case II example into a Case I example as a Horn situation. In this extended sense, there are three types of situations which can develop into Horn situations. For other situations, or other processes we get counter examples for Horn's division of pragmatic labour.

## 3    Conclusions

The diachronic learning model allows only to calculate the associated meaning for each form separately. In comparison, the approach using diachronic laws is much easier to handle and it allows characterising global rules for meaning shifts more straightforwardly. E.g. we did show how and when Horn situations can emerge out of Blutner-squares. Compared to Horn's principle and the principle of weak optimality the diachronic laws have an additional empirical justification because they are derived from an underlying learning theory. They avoid the problem of over–generation and lead to different empirical predictions for when meaning shifts can occur and when not.

## References

[1] A. Benz (2003): *Partial Blocking, Associative Learning, and the Principle of Weak Optimality;* In: J. Spenader, A. Eriksson, Ö. Dahl (eds.): *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, pp. 150-159; Stockholm.

[2] R. Blutner (2000): *Some Aspects of Optimality in Natural Language Interpretation;* In: Helen de Hoop & Henriette de Swart (eds.) *Papers on Optimality Theoretic Semantics*. Utrecht Institute of Linguistics OTS, December 1999, pp 1-21. Also: *Journal of Semantics* 17, pp. 189-216.

[3] L. Horn (1984): *Towards a new taxonomy of pragmatic inference: Q–based and R–based implicature;* In: D. Schiffrin (ed.): *Meaning, Form, and Use in Context: Linguistic Applications*, Georgetown University Press, Washington, pp. 11–42.

[4] J. Mattausch (November 2000): *On Optimization in Discourse Generation;* master thesis, Universiteit van Amsterdam.

[5] T. Mitchell (1997): *Machine Learning;* McGraw-Hill, New York.