Tolerance and mixed consequence in a super/sub-valuationist setting

Pablo Cobreros
Paul Egre
Dave Ripley
Robert van Rooij

September 4, 2010

Abstract

In a previous paper (see 'Tolerant, Classical, Strict', henceforth TCS) we investigated a semantic framework to deal with the idea that vague predicates are tolerant, namely that small changes do not affect the applicability of a vague predicate even if large changes do. Our approach there rests on two main ideas. First, given a classical extension of a predicate, we might define a *strict* and a *tolerant* extension depending on an indifference relation associated to that predicate. Second, we can use these notions of satisfaction to lead to *mixed* consequence relations that capture non-transitive tolerant reasoning. The present paper intends to explore the possibility of defining mixed notions of consequence in a super/sub-valuationist setting and see to what extent any of these notions captures non-transitive tolerant reasoning.

1 Introduction

Take a (long enough) series $\langle a_1, a_2 \dots a_n \rangle$ of patches of color. The first is clearly red and the last is clearly orange (and so, not red). However, each patch in the series is only imperceptibly different in color from its successor, so that an indifference relation holds between any adjacent patches in the series. This relation is, we take it, reflexive, symmetric but not transitive. Vague expressions such as 'red' seem to be tolerant in the sense that a small enough difference in the color of the patches cannot affect the applicability of the predicate, even if big enough differences do. In this sense, if we take any two adjacent patches from our series a and b, it seems that from the fact that a is red we can confidently conclude that b is red as well. We can construct the following step-by-step sorites argument:

a_1 is red
a_1 is only imperceptibly different from a_2
Therefore: a_2 is red

a_2 is red
a_2 is only imperceptibly different from a_3
Therefore: a_3 is red

$\vdots$

a_{n-1} is red
a_{n-1} is only imperceptibly different from a_n
Therefore: a_n is red

These arguments, however, can be classically joined together so that the following argument is also valid:

a_1 is red
a_1 is only imperceptibly different from a_2
a_2 is only imperceptibly different from a_3

$\vdots$

a_{n-1} is only imperceptibly different from a_n
Therefore: a_n is red

Another way to look at the tolerance of vague predicates is by directly considering a formulation of the tolerance principle:

(T) $\forall x \forall y(P(x) \land xI_Py \rightarrow P(y))$, where I_P expresses the P-relevant indifference relation.

However, the tolerance principle classically entails that all members of the series are red, contradicting the fact that the last is orange. Due to this fact and the previous soritical argument, an important group of solutions to the sorites paradox consists in rejecting the tolerance principle along with tolerant reasoning broadly considered. For example, for the epistemicist the tolerance principle is false and, in fact, there is a last item in our series that is red followed by a non-red item. For some philosophers endorsing a many-valued semantics (such as Kleene's strong three-valued logic, K3) the tolerance principle is not true (though not false either); for others, such as supervaluationists, the tolerance principle is in fact false although there is no falsifying instance. Further, K3 semantics and supervaluationist semantics do not allow for tolerant reasoning since it is not the case that from the fact that a and b are similar enough in P-relevant respects and a is P, it logically follows that b is P (since a might be P-similar to b and truly P while b is not truly P).

In TCS we develop a semantic framework originally proposed by van Rooij (2010) in order to accommodate the idea that vague predicates are tolerant. The basic idea is that the semantics of a vague predicate P can be made sensitive to the P-relevant indifference relation. A tolerant first-order model $\langle D, I, \sim \rangle$, is a classical first-order model $\langle D, I \rangle$ expanded with a function $\sim$ that maps every predicate of the language to a binary relation $\sim_P$ that is symmetric and reflexive (but possibly non-transitive). Classical satisfaction in a tolerant model $\langle D, I, \sim \rangle$ is defined as classical satisfaction in $\langle D, I \rangle$; classical validity in a tolerant model is defined accordingly as classical satisfaction in every tolerant model. Now we define, for tolerant models, the dual notions of *tolerant* and *strict* satisfaction making use of classical satisfaction and indifference relations. A sentence Pa is tolerantly true in a tolerant model M (in symbols: $M \vDash^t Pa$) iff there is an x such that $a \sim_P x$ and Px is classically true; a sentence Pa is strictly true in a model M (in symbols: $M \vDash^s Pa$) iff for every x if $a \sim_P x$ then Px is classically true. We extend the notion of tolerant and strict satisfaction to arbitrary formulae by simultaneous induction; in particular $M \vDash^t \neg\varphi$ iff $M \nvDash^s \varphi$ and $M \vDash^s \neg\varphi$ iff $M \nvDash^t \varphi$ (so that '$\vDash^t$' and '$\vDash^s$' are duals).[1]

An interesting feature of this semantics is that the tolerance principle (T) is tolerantly valid. In TCS we show that the logics obtained by defining logical consequence as preservation of strict truth and preservation of tolerant truth, coincide (for the classical vocabulary without identity), respectively, with strong Kleene logic (K3) and its dual, Priest's Logic of Paradox (LP). Though we might endorse the tolerance principle given tolerant satisfaction, we argue in TCS that the logic resulting from the definition of logical consequence as preservation of tolerance truth (that is, LP) does not provide an adequate framework; in particular, *modus ponens* is not an LP-valid rule of inference. So we consider the notions of logical consequence resulting of mixing any of our three notions of satisfaction. It turns out that the notion of logical consequence that goes from strictly true premises to a tolerantly true conclusion (we call it *st*-entailment: $\vDash^{st}$) leads to a non-transitive logic in which the tolerance principle is valid and where both *modus ponens* and the deduction theorem hold. This notion of logical consequence, we take it, provides a nice framework in which we can provide a tolerant solution to the paradox. Let I_P express in the language the similarity relation $\sim_P$. We have that $Pa, aI_Pb \vDash^{st} Pb$, but $Pa_1, \forall x a_i I_P a_{i+1} \nvDash^{st} Pa_n$ (for $1 \leq i < n$) for a large enough n. So, although each step in the argument is *st*-valid, the result of joining all the steps together is not *st*-valid.

As mentioned above, tolerant and strict satisfaction lead, when we consider unmixed consequence, to K3 and LP respectively. These logics bear an analogy to supervaluationist and subvaluationist logics: both K3 and

[1] See TCS sec. 1.4 for a full description of the semantics.

supervaluationist logic are *paracomplete*, and LP and subvaluationist logic are their *paraconsistent* duals.[2] Interestingly, supervaluationism (subvaluationism) is *weakly* paracomplete (paraconsistent) in the sense that classical validities (contradictions) are supervaluationistically valid (subvaluationistically unsatisfiable). The aim of the present work is to extend the research on tolerance and mixed consequence to the super- and subvaluationist setting (the s'valuationist setting, for short) in order to compare how many of our previous results can be transposed to this new framework. More specifically, the work aims to address the following questions:

1. What are the relations between the different notions of logical consequence (pure or mixed) that we might define in an s'valuationist setting? (that is, the notions of logical consequence that we might define out of supertruth, subtruth and a suitable analogue of classical truth)

2. How can we connect indifference relations s'valuationist semantics and to what extent can we use this connection to provide a tolerant solution to the sorites paradox?

3. What are the similarities/differences between this and our previous approach? Is there any definitive advantage of one approach over the other?

The present discussion is concerned with the language of first-order logic. For simplicity we will focus on languages with just monadic predicates, constants and without identity or other polyadic predicates. We aim to compare the different logics with respect to different languages. Our "restricted vocabulary" is an ordinary first order language (again, without identity or other polyadic predicates); our "full vocabulary" includes in addition a binary similarity predicate I_P for each monadic predicate P in the language. The predicates I_P will express the similarity relations $\sim_P$; as we will see, the relations between various notions of consequence are sensitive to the presence or absence of these I_P predicates.[3] Many results in TCS transpose to the s'valuationist setting; in order to appreciate this fact (but also to see when new twists occur) in brackets we make systematic cross-reference to the corresponding results in our previous paper.

[2] A consequence relation $\vDash^x$ is paracomplete iff there are A, B such that $B \vDash^x \{A, \neg A\}$ does not hold, and paraconsistent iff there are A, B such that $\{A, \neg A\} \vDash^x B$ does not hold.

[3] After Lemma 2 below we make use of a modality to illustrate a small remark concerning this lemma; however, we do not consider modalities as part of the full vocabulary, at least in this paper.

The structure of the paper is as follows. Section 2 briefly introduces s'valuationist models (S-models) along with three notions of satisfaction: supertruth, local truth and subtruth. Section 3 deals with mixed consequence in a s'valuationist setting. In the first place we characterize the relations between the different logics for the restricted vocabulary (subsection 3.1). Then we propose how to interpret similarity relations in the present framework and spell out the relations between logics for the full vocabulary (subsection 3.2). We close our discussion by briefly considering tolerance and the sorites in the present framework and a comparison with our previous proposal (subsection 3.3). The appendix provides a tableau-based system to check for any of the notions of logical consequence discussed in this paper.

2 Supertruth, local truth and subtruth

Supervaluationism and subvaluationism agree on the idea that a vague expression can be made precise in several ways consistent with the use we make of it. These theories disagree, however, on what it takes for a sentence to be true. An admissible precisification is a classical model respecting some constraints depending on the meaning of expressions, like analytic relations between expressions (nothing is counted both as a child and as a baby) and comparative relations (nothing taller than x is counted as not tall in a precisification where x is counted as tall). According to supervaluationism a sentence is true (supertrue) just in case it is true in *every* admissible precisification; thus vagueness amounts to some form of *underdetermination* of meaning. According to subvaluationism a sentence is true (subtrue) just in case it is true in *some* admissible precisification; thus vagueness amounts to some form of *overdetermination* of meaning. It is clear from the previous informal remarks that we can construct s'valuationist models out of classical models:

A classical model is a tuple $M = \langle D, I \rangle$ such that:

- D is a non-empty domain of individuals and
- I is an interpretation function for the non-logical vocabulary mapping constants to individuals in D and predicates to subsets of D.

Following a standard definition of classical satisfaction, we write $M \vDash \varphi$ to mean that φ is classically true in M.

An s'valuationist model $\mathscr{M}$ is a non-empty set of admissible classical models[4] where for any two models $M = \langle D, I\rangle \in \mathscr{M}$ and $M' = \langle D', I'\rangle \in \mathscr{M}$:

- $D = D'$ and
- $I(c) = I'(c)$ for every constant c.

In order to define an analogue of classical satisfaction we will further consider "S-models" $\langle \mathscr{M}, M\rangle$ in which $M \in \mathscr{M}$ (M can be thought of as a kind of designated model in $\mathscr{M}$). It is convenient for reasons of notation to use the index p for su*p*ertruth and b for su*b*truth even if this is not standard usage.

Definition 1.
Supertruth*: A sentence φ is supertrue in an S-model $\langle \mathscr{M}, M\rangle$, (written $\mathscr{M}, M \vDash^p \varphi$) iff for all $M' \in \mathscr{M}$, $M' \vDash \varphi$.*
Local truth*: φ is locally true in an S-model $\langle \mathscr{M}, M\rangle$, (written $\mathscr{M}, M \vDash^l \varphi$) iff $M \vDash \varphi$.*
Subtruth*: A sentence φ is subtrue in an S-model $\langle \mathscr{M}, M\rangle$, (written $\mathscr{M}, M \vDash^b \varphi$) iff for some $M' \in \mathscr{M}$, $M' \vDash \varphi$.*

These notions of satisfaction resemble our previous notions of strict, classical and tolerant satisfaction respectively. In the first place, $\vDash^l$ is self-dual in the sense that for any sentence and S-model $\mathscr{M}, M \vDash^l \varphi$ iff $\mathscr{M}, M \nvDash^l \neg\varphi$ while $\vDash^p$ and $\vDash^b$ are duals since $\mathscr{M}, M \vDash^p \varphi$ iff $\mathscr{M}, M \nvDash^b \neg\varphi$ (and $\mathscr{M}, M \vDash^b \varphi$ iff $\mathscr{M}, M \nvDash^p \neg\varphi$). In the second place, each notion sets different standards for satisfaction. It is *harder* for a sentence to be supertrue in an S-model than to be locally true and it is *harder* to be locally true in an S-model than to be subtrue, as is stated in the following easy lemma:

Lemma 1 (Compare TCS Lemma 1)**.** *For any S-model $\langle \mathscr{M}, M\rangle$ and any sentence φ, $\mathscr{M}, M \vDash^p \varphi \Rightarrow \mathscr{M}, M \vDash^l \varphi \Rightarrow \mathscr{M}, M \vDash^b \varphi$*

Proof. If every φ is true in every model in $\mathscr{M}$, then it is certainly true in M and so $\mathscr{M}, M \vDash^l \varphi$. In turn, if φ is true in M then certainly there is at least an M' in $\mathscr{M}$ at which φ is true. $\square$

A *vague interpretation* (in the present context) is an interpretation where some sentences are neither supertrue nor superfalse; equivalently, a vague interpretation is an interpretation where some sentences are both subtrue and subfalse. Thus, a vague interpretation is a supermodel $\mathscr{M}$ that contains

[4] Fine's presentation of supervaluationism (Fine (1975)) is different from the present one in that he starts out from a partial model and defines a supermodel as a set containing that model and all its classical extensions. However, when one considers *admissibility constraints* the partial model drops out of the picture (Kremer and Kremer, 2003, 234).

at least two distinct classical models (two models that disagree in the interpretation of some of the predicates). Accordingly, a *precise interpretation* is a supermodel containing just one classical model. Naturally, local truth, supertruth and subtruth coincide for precise interpretations.

Since the restricted vocabulary cannot *see* what is going on in models different from the designated model, any S-model can be reduced to a precise S-model which is equivalent over the restricted vocabulary with respect to local satisfaction; in this new S-model, in turn, local truth, supertruth and subtruth coincide.

Lemma 2 (Compare TCS Lemma 2)**.** *Let $\langle \mathscr{M}, M\rangle$ be an S-model. Let $\langle \mathscr{M}', M\rangle$ be the model obtained from $\langle \mathscr{M}, M\rangle$ and taking M as the sole model in $\mathscr{M}$. Then for every sentence φ in the restricted vocabulary, $\mathscr{M}, M \vDash^l \varphi$ iff $\mathscr{M}', M \vDash^l \varphi$ iff $\mathscr{M}', M \vDash^p \varphi$ iff $\mathscr{M}', M \vDash^b \varphi$.*

Proof Sketch. By induction, in the restricted vocabulary, whether $\mathscr{M}, M \vDash^l \varphi$ depends just on the model M in $\mathscr{M}$. Thus, S-models $\langle \mathscr{M}, M\rangle$ and $\langle \mathscr{M}', M\rangle$ are $\vDash^l$-equivalents. Since $\mathscr{M}', M$ contains a single model, any sentence φ will be true in every model, just in case it is true in some model, just in case it is true in M. $\square$

The lemma is clearly linked to the restricted vocabulary. If we allow expressions that *can see* what is going on in other models, S-models $\langle \mathscr{M}, M\rangle$ and $\langle \mathscr{M}', M\rangle$ might cease to be $\vDash^l$-equivalents. For example, define for any S-model $\langle \mathscr{M}, M\rangle$: $\mathscr{M}, M \vDash^l \Box\varphi$ just in case for all $M^* \in \mathscr{M}$, $M^* \vDash \varphi$. The sentence $\varphi \wedge \neg\Box\varphi$ will be true in some S-models $\langle \mathscr{M}, M\rangle$, but false in any precise model.

3 Mixed consequence

The consequence relation corresponding to preservation of local truth in every model is, in the restricted vocabulary, classical logic. In turn, preservation of supertruth and preservation of subtruth lead to supervaluationist logic and subvaluationist logic respectively. However, in addition to *pure* forms of logical consequence, we might consider the notions of consequence resulting of mixing different notions of satisfaction. In section 3.1 we study the relation between possible combinations of logical consequence for the restricted vocabulary. In section 3.2 we introduce similarity relations and work out the relations between the different logics for a language containing similarity predicates. First of all, a structured way to talk about these consequence relations:

Definition 2. $\Gamma \vDash^{mn} \Delta$ just in case for every S-model $\mathscr{M}, M$: if $\forall \gamma \in \Gamma, \mathscr{M}, M \vDash^{m} \gamma$ then $\exists \delta \in \Delta, \mathscr{M}, M \vDash^{n} \delta$.

So, for example, $\vDash^{pp}$ is supervaluationist consequence, $\vDash^{bb}$ is subvaluationist consequence and $\vDash^{ll}$ is classical consequence (at least for the restricted vocabulary). However, we can also consider *mixed* versions as, for example, $\vDash^{pb}$ that preserves a subtrue conclusion from supertrue premises.

As pointed out before, $\vDash^{p}$ and $\vDash^{b}$ are dual notions of satisfaction while $\vDash^{l}$ is self-dual. We define now more generally the notion of dual for consequence relations and point out this relation between our nine notions of logical consequence.

Definition 3 (Dual consequence relation)**.** Let $\vDash^{x}$ be a notion of logical consequence. Its *dual* is the notion of logical consequence $\vDash^{y}$ such that: $\Gamma \vDash^{x} \Delta$ iff $\neg(\Delta) \vDash^{y} \neg(\Gamma)$ (where $\neg(\Delta) = \{\neg\delta \mid \delta \in \Delta\}$)

These are the resulting duality relations:

1. $\vDash^{ll}$, $\vDash^{pb}$ and $\vDash^{bp}$ are self-dual.
2. $\vDash^{pp}$ and $\vDash^{bb}$ are duals.
3. $\vDash^{pl}$ and $\vDash^{lb}$ are duals.
4. $\vDash^{lp}$ and $\vDash^{bl}$ are duals.

3.1 The restricted vocabulary

Lemma 3 (Compare TCS Lemma 7)**.** *For any m: $\vDash^{bm} \subseteq \vDash^{lm} \subseteq \vDash^{pm}$ and $\vDash^{mp} \subseteq \vDash^{ml} \subseteq \vDash^{mb}$.*

Proof. Since we know that, for any S-model $\langle \mathscr{M}, M \rangle$, $\{\varphi : \mathscr{M}, M \vDash^{p} \varphi\} \subseteq \{\varphi : \mathscr{M}, M \vDash^{l} \varphi\} \subseteq \{\varphi : \mathscr{M}, M \vDash^{p} \varphi\}$ (Lemma 1), it follows that if an S-model $\langle \mathscr{M}, M \rangle$ is a pm-counterexample to an argument, it is also an lm-counterexample, and if it is an lm-counterexample, it is also a bm-counterexample. Similarly, if a model is an mb-counterexample to an argument, it must also be an ml-counterexample, and if it is an ml-counterexample, it must also be an mp-counterexample. $\square$

The lemma is based directly on the definitions of satisfaction and it holds for the full vocabulary as well. This lemma answers some questions regarding the relation between our nine notions of consequence. We complete the picture for the restricted vocabulary.

a) $\vDash^{ll} = \vDash^{pl} = \vDash^{lb} = \vDash^{pb}$

Lemma 4 (Compare TCS Lemma 8)**.** $\Gamma \vDash^{pb} \Delta \Rightarrow \Gamma \vDash^{ll} \Delta$

Proof. Assume $\Gamma \nvDash^{ll} \Delta$, then:
There is an $\langle \mathscr{M}, M\rangle$ s. t. $\forall\gamma \in \Gamma\mathscr{M}, M \vDash^{l} \gamma$ and $\forall\delta \in \Delta\mathscr{M}, M \nvDash^{l} \delta$

$$\Downarrow \qquad \text{(by Lemma 2)}$$

There is an $\langle \mathscr{M}', M\rangle$ s. t. $\forall\gamma \in \Gamma\mathscr{M}', M \vDash^{p} \gamma$ and $\forall\delta \in \Delta\mathscr{M}', M \nvDash^{b} \delta$ □

Lemma 3 tells us that $\vDash^{ll}\subseteq\vDash^{pb}$ and so $\vDash^{ll}=\vDash^{pb}$. The same lemma states that $\vDash^{ll}\subseteq\vDash^{pl}\subseteq\vDash^{pb}$ and $\vDash^{ll}\subseteq\vDash^{lb}\subseteq\vDash^{pb}$ so $\vDash^{ll}=\vDash^{pl}=\vDash^{lb}=\vDash^{pb}$.

b) $\vDash^{pp}$ and $\vDash^{bb}$ are distinct and strictly weaker than $\vDash^{ll}$

Note that $\emptyset \nvDash^{pp} \{p, \neg p\}$ but $\emptyset \vDash^{bb} \{p, \neg p\}$ and, dually, $\{p, \neg p\} \nvDash^{bb} \emptyset$ but $\{p, \neg p\} \vDash^{pp} \emptyset$. So neither consequence relation contains the other. Now $\{p, \neg p\} \vDash^{ll} \emptyset$ and $\emptyset \vDash^{ll} \{p, \neg p\}$ and so if $\vDash^{pp}$ and $\vDash^{bb}$ are both weaker than $\vDash^{ll}$, they are strictly weaker.

Lemma 5. $\Gamma \vDash^{pp} \Delta \Rightarrow \Gamma \vDash^{ll} \Delta$ *and* $\Gamma \vDash^{bb} \Delta \Rightarrow \Gamma \vDash^{ll} \Delta$

Proof. Assume $\Gamma \nvDash^{ll} \Delta$, then:
There is an $\langle \mathscr{M}, M\rangle$ s. t. $\forall\gamma \in \Gamma\mathscr{M}, M \vDash^{l} \gamma$ and $\forall\delta \in \Delta\mathscr{M}, M \nvDash^{l} \delta$

$$\Downarrow \qquad \text{(by Lemma 2)}$$

There is an $\langle \mathscr{M}', M\rangle$ s. t. $\forall\gamma \in \Gamma\mathscr{M}', M \vDash^{p} \gamma$ and $\forall\delta \in \Delta\mathscr{M}', M \nvDash^{p} \delta$

(and similarly for the second claim) □

c) $\vDash^{lp}$ is strictly weaker than $\vDash^{pp}$ and $\vDash^{bl}$ is strictly weaker than $\vDash^{bb}$

From Lemma 3 we have that $\vDash^{bl}\subseteq\vDash^{bb}$ and $\vDash^{lp}\subseteq\vDash^{pp}$. Unlike $\vDash^{pp}$ and $\vDash^{bb}$, however, $\vDash^{bl}$ or $\vDash^{lp}$ are not reflexive (since a formula might be subtrue in a model without being locally true in that model, and similarly for $\vDash^{lp}$).

d) $\vDash^{bp}$ is strictly weaker than both $\vDash^{bl}$ and $\vDash^{lp}$

From Lemma 3 we have that $\vDash^{bp}\subseteq\vDash^{bl}$ and $\vDash^{bp}\subseteq\vDash^{lp}$. To see that the inclusion is strict notice that $\emptyset \vDash^{bl} \{p, \neg p\}$ but $\emptyset \nvDash^{bp} \{p, \neg p\}$ and $\{p, \neg p\} \vDash^{lp} \emptyset$ but $\{p, \neg p\} \nvDash^{bp} \emptyset$. However, $\vDash^{bp}$ is not the empty relation, since, for example, $\emptyset \vDash^{bp} \{p \wedge \neg p\}$. In fact, $\vDash^{bp}$ is the weakest consequence relation preserving the validity of classical tautologies and the unsatisfiability of classical contradictions:

Lemma 6 (Compare TCS Lemma 9)**.** $\Gamma \vDash^{bp} \Delta$ *iff either* $\Gamma \vDash^{bp} \emptyset$ *or* $\emptyset \vDash^{bp} \Delta$.

Proof. For the right to left direction note that if $\Gamma \vDash^{bp} \emptyset$ then $\Gamma \vDash^{bp} \Delta$ and if $\emptyset \vDash^{bp} \Delta$ then $\Gamma \vDash^{bp} \Delta$.

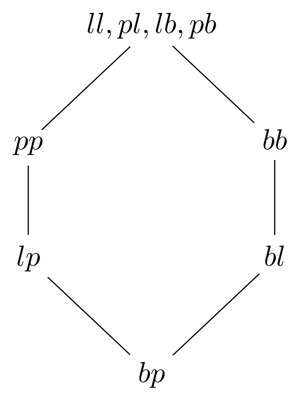
For the left to right direction, assume that $\Gamma \nvDash^{bp} \emptyset$ and $\emptyset \nvDash^{bp} \Delta$ for some Γ and Δ. If both Γ and Δ are empty then it is clear that $\Gamma \nvDash^{bp} \Delta$. So assume that at least one is non-empty. Construct an S-model $\langle \mathscr{M}, M \rangle$ following these rules:

- For each $\gamma \in \Gamma$ include a classical model M_γ in $\mathscr{M}$ such that $M_\gamma \vDash \gamma$ (the fact that $\Gamma \nvDash^{bp} \emptyset$ guarantees that there is such a model).
- For each $\delta \in \Delta$ include a classical model M_δ in $\mathscr{M}$ such that $M_\delta \nvDash \delta$ (the fact that $\emptyset \nvDash^{bp} \Delta$ guarantees that there is such a model).
- The "designated" model M in $\mathscr{M}, M$ can be any model in $\mathscr{M}$ (that there is some such model is guaranteed by the assumption that at least one of Γ and Δ is non-empty).

The model shows that $\Gamma \nvDash^{bp} \Delta$. $\square$

Summing up

On the restricted vocabulary four of our nine notions of logical consequence collapse and so there are six different notions of logical consequence. The strongest consequence relation is $\vDash^{ll}$ (which in the restricted vocabulary is just classical consequence) which turns out to be equivalent to $\vDash^{lb}$, $\vDash^{pl}$ and $\vDash^{bp}$. $\vDash^{pp}$ and $\vDash^{bb}$ (super- and subvaluationist consequence) are distinct and strictly weaker than $\vDash^{ll}$. $\vDash^{lp}$ is strictly weaker than $\vDash^{pp}$ and $\vDash^{bl}$ strictly weaker than $\vDash^{bb}$. Finally, $\vDash^{bp}$ is strictly weaker than any of the other relations. The picture is, thus, as follows:



3.2 Similarity relations

In this section we want to focus on models including a similarity relation for each predicate P in the language and see how the presence of these relations should be reflected in the semantics.

An ST-model is a triple $\langle \mathscr{M}, M, \sim \rangle$ where $\mathscr{M}$ and M are as before and $\sim$ is a function mapping each predicate P of the language to a relation $\sim_P$ in $D \times D$ that is reflexive, symmetric but possibly non-transitive. The definitions of $\vDash^l$, $\vDash^p$ and $\vDash^b$ carry over from S-models to ST-models; similarly for validity and logical consequence. Similarity relations will be crisply interpreted in ST-models in the sense that for any individuals a and b, similarity predicate I_P and ST-model $\langle \mathscr{M}, M, \sim \rangle$: $\mathscr{M}, M \vDash^l aI_Pb$ iff $\mathscr{M}, M \vDash^p aI_Pb$ iff $\mathscr{M}, M \vDash^b aI_Pb$ iff $a \sim_P b$.[5]

The introduction of similarity relations should be reflected in the semantics imposing a constraint on models. A natural idea is that any ST-model in which $a \sim_P b$ is such that if Pa locally holds, then Pb holds at least subvaluationally and if $\neg Pa$ locally holds then $\neg Pb$ holds, at least subvaluationally.

Given an ST-model $\langle \mathscr{M}, M, \sim \rangle$, any individuals a and b in the model, and any predicate P, if $a \sim_P b$ then $\exists M' \in \mathscr{M}$ s. t. $M \vDash Pa$ iff $M' \vDash Pb$.

The motivation for this constraint is that sentences Pa and Pb cannot have a big difference in semantic status if a and b are P-similar.

Given the previous characterization of similarity relations, if we allow similarity predicates in the language then *pb*, *pl* and *lb* are stronger than *ll* since the inference from $\{Pa, aI_Pb\}$ to Pb is not *ll*-valid, but it is *pb*, *pl* and *lb* valid (the inference is neither *pp* or *bb* valid). And *pb* is stronger than both *pl* and *lb* since the inference from $\{Pa, aI_Pb, bI_Pc\}$ to Pc is *pb* but not *pl* and *lb* valid (*pb* allows us to take two steps in the sorites series but not more). The consequence relations *pl* and *lb* come apart since, on the one hand $\{Pa, aI_Pb, \neg Pb\} \vDash^{pl} \emptyset$ while $\{Pa, aI_Pb, \neg Pb\} \nvDash^{lb} \emptyset$ and on the other hand $\emptyset \vDash^{lb} \{Pa, \neg aI_Pb, \neg Pb\}$ while $\emptyset \nvDash^{pl} \{Pa, \neg aI_Pb, \neg Pb\}$. Thus, this is the picture so far:

[5] Here we follow the same strategy as in TCS (p. 5-6). The non-crisp interpretation of similarity predicates is relevant for the problem of higher-order vagueness; however we leave this issue for future discussion.

$$\begin{tikzcd}
& pb \arrow[dl, dash] \arrow[dr, dash] & \\
pl \arrow[dr, dash] & & lb \arrow[dl, dash] \\
& ll &
\end{tikzcd}$$

The next question is whether pp and bb are still weaker than ll. The answer is negative since $\{Pa, aI_Pb, \neg Pb\} \vDash^{pp} \emptyset$ and $\emptyset \vDash^{bb} \{Pa, \neg aI_Pb, \neg Pb\}$ but neither inference hold in the case of ll. Given Lemma 3 again, lp is still weaker than both pp and ll, bl is weaker than bb and ll and bp is weaker than lp and bl. Since the examples showing distinctness in section 3.2 still work, these inclusions are strict:

$$\begin{tikzcd}
& pb \arrow[dl, dash] \arrow[dr, dash] & \\
pl \arrow[d, dash] \arrow[dr, dash] & & lb \arrow[dl, dash] \arrow[d, dash] \\
pp \arrow[d, dash] & ll \arrow[dl, dash] \arrow[dr, dash] & bb \arrow[d, dash] \\
lp \arrow[dr, dash] & & bl \arrow[dl, dash] \\
& bp &
\end{tikzcd}$$

Since *modus ponens* is ll-valid it is valid in any of the three stronger notions of consequence. However, the deduction theorem (for finite Γ and Δ, $\Gamma \vDash^{mn} \Delta$ iff $\vDash^{mn} \bigwedge \Gamma \to \bigvee \Delta$) does not hold for all the nine consequence relations. For example, it does not hold for bp even in the restricted vocabulary since $\vDash^{bp} \varphi \to (\varphi \vee \neg\varphi)$ though $\varphi \nvDash^{bp} \varphi, \neg\varphi$. It does not hold in the expanded vocabulary for some other consequence relations. For example, $\{Pa, aI_Pb, \neg Pb\} \vDash^{pl} Pc$ but $\nvDash^{pl} (Pa \wedge aI_Pb \wedge \neg Pb) \to Pc$.

In TCS we provide a result linking the deduction theorem with self-duality (p. 23, Lemma 10). That result, however, cannot be fully transposed to the present setting; the reason is that the present setting does not always preserve the standard connection between the comma in the premises and

'$\wedge$' on the one hand, and the comma in the conclusions and '$\vee$' on the other (as evidenced by the fact that $\{p, \neg p\}$ is b-satisfiable though $p \wedge \neg p$ is not and the fact that $p \vee \neg p$ is p-valid though $\{p, \neg p\}$ is not). Nevertheless, the deduction theorem still holds for pb:

Lemma 7. *For finite Γ and Δ, $\Gamma \vDash^{pb} \Delta$ iff $\vDash^{pb} \bigwedge \Gamma \to \bigvee \Delta$*

Proof. Assume: $\Gamma \vDash^{pb} \Delta$ iff

1. For any ST-model $\langle \mathscr{M}, M \rangle$:
   either $\exists \gamma \in \Gamma\ \mathscr{M}, M \nvDash^{p} \gamma$ or $\exists \delta \in \Delta\ \mathscr{M}, M \vDash^{b} \delta$ iff
2. For any ST-model $\langle \mathscr{M}, M \rangle$:
   either $\mathscr{M}, M \nvDash^{p} \bigwedge \Gamma$ or $\mathscr{M}, M \vDash^{b} \bigvee \Delta$ iff
3. For any ST-model $\langle \mathscr{M}, M \rangle$:
   either $\mathscr{M}, M \vDash^{b} \neg \bigwedge \Gamma$ or $\mathscr{M}, M \vDash^{b} \bigvee \Delta$ iff
4. For any ST-model $\langle \mathscr{M}, M \rangle$: $\mathscr{M}, M \vDash^{b} \bigwedge \Gamma \to \bigvee \Delta$ iff
5. $\vDash^{pb} \bigwedge \Gamma \to \bigvee \Delta$

□

Step from 1 to to 2 is based on the fact that a conjunction fails to be supertrue iff some conjunct fails to be supertrue. Similarly, a disjunction is subtrue iff some disjunct is subtrue. Step from 2 to 3 is based on the duality of $\vDash^{p}$ and $\vDash^{b}$. Step from 3 to 4 is also based on the fact that $\vDash^{b} \varphi$ or $\vDash^{b} \psi$ iff $\vDash^{b} \varphi \vee \psi$.

3.3 Tolerance

3.3.1 Non-transitive reasoning and the sorites

We repeat the sorites argument presented in section 1. Take a series of patches of color: $\langle a_1, a_2 \ldots a_n \rangle$. The first is clearly red and the last is clearly orange (and so clearly not red). Each pair of adjacent members of the series is similar in P-relevant respects; that is, $a_i \sim_P a_{i+1}$ for $1 \leq i < n$. We can construct the following sorites argument:

a_1 is red
a_1 is only imperceptibly different from a_2
Therefore: a_2 is red

a_2 is red
a_2 is only imperceptibly different from a_3
Therefore: a_3 is red

$\vdots$

a_{n-1} is red
a_{n-1} is only imperceptibly different from a_n
Therefore: a_n is red

As pointed out before, for any particular a and b, the inference from Pa, aI_Pb to Pb is valid for *pb*, *pl* and *lb*. So each step of the argument is valid according to any of these notions of consequence. However, the argument,

a_1 is red
a_1 is only imperceptibly different from a_2
a_2 is only imperceptibly different from a_3

$\vdots$

a_{n-1} is only imperceptibly different from a_n
Therefore: a_n is red

Is no longer valid in any of the three notions of consequence. In effect, these notions of consequence allow us to solve this version of the paradox as a case of non-transitive reasoning. We take it, however, that among the three notions of consequence, *pb* is the best candidate for an adequate characterization of non-transitive reasoning since it is self-dual (which preserves standard relations between validity and unsatisfiability) and, further, it preserves not only the validity of *modus ponens* but also the deduction theorem.

The notion of consequence *pb*, however, makes a different diagnosis of the problem with sorites arguments when we consider the formulation involving the tolerance principle,

(T) $\forall x \forall y (P(x) \land xI_P y \to P(y))$

Each instance of the tolerance principle is *pb*-valid and, correspondingly, the negation of any instance is *pb*-unsatisfiable. However, the situation changes when we look at the tolerance principle itself: it is not valid even in our most generous notion of satisfaction $\vDash^b$. This fact is linked to the classicality characteristic of the notions of supertruth and superfalsity. In effect, as is well known, the tolerance principle is classically false for any suitable sorites series, which makes its negation true in any classical model respecting the constraints of any suitable sorites series. This makes the principle superfalse (and its negation supertrue) in any suitable S-model.

So the situation regarding the solution to the sorites is different depending on the formulation of the paradox. For the formulation involving similarity relations in a chain of step-by-step arguments the diagnosis is that each step is valid, but not the corresponding 'conjoined' argument. For the formulation involving (T) the diagnosis is that, although each instance is valid, the principle itself is not valid; further, the principle is superfalse in any suitable S-model; thus, in this formulation the argument is unsound. In the next and last section we briefly compare this situation with the situation in our previous paper concerning our preferred notion of logical consequence '*st*'.

3.3.2 Comparisons

As pointed out in section 1, in TCS we develop a semantics for vague predicates that is sensitive to indifference relations. We define two notions of satisfaction that play a central role: strict and tolerant satisfaction. The first leads to a notion of *unmixed* consequence equivalent (in the restricted vocabulary) to strong Kleene logic K3; the second to a notion of *unmixed* consequence equivalent (in the restricted vocabulary) to K3's dual: Priest's Logic of Paradox LP (sect 2.2. TCS). As pointed out before, K3 is a paracomplete logic and LP is its paraconsistent dual. The notion of logical consequence to which we give more credit in TCS is *st*: the logic going from strictly true premises to tolerantly true conclusions.

Here, we focused on *pb*, which is built on the supervaluationist and subvaluationist notions of satisfaction. As in the case of K3, the supervaluationist logic is a paracomplete logic; as in the case of LP, the subvaluationist's logic is a paraconsistent logic. However, unlike K3, the supervaluationist's logic is only *weakly* paracomplete in the sense that classical validities remain valid. In the same vein, unlike LP, the subvaluationist's logic is only *weakly* paraconsistent in the sense that classical contradictions remain unsatisfiable (K3 and LP are accordingly called 'strongly' paracomplete/paraconsistent).[6]

For sorites formulations involving similarity relations in a chain of arguments (let us call them 'type A' arguments (see TCS version 1 argument)), *st* and *pb* make similar predictions; namely, the argument is invalid. This is natural since the prediction is based on similar features of *st* and *pb*, namely, going from high-standards of satisfaction in the premises to lower standards in the conclusions with suitable constraints on similarity predicates. The difference comes when we consider the formulation of the paradox involving

[6]See (Hyde, 2008, ch. 4) for discussion on this distinction in the context of a theory of vagueness. Hyde credits the distinction to Arruda (1989).

the tolerance principle (call these 'type B' arguments (see TCS, version 2 argument)). Both *st* and *pb* agree that the argument is unsound; however, whereas for *st* the tolerance principle is valid (as is any instance of it), for *pb*, although each instance is valid, the principle itself is not valid (is not even subtrue in any suitable model).

	st-consequence	*pb*-consequence
(T)	valid	not valid
Instances	valid	valid
Diagnosis	The argument is unsound	The argument is unsound

Figure 1: *st*, *pb* and B-arguments

The situation of B-type arguments reflects the nature of the notions of satisfaction involved in each notion of consequence. On the one hand, it is characteristic of paracomplete solutions to the paradox to diagnose sorites arguments as valid but unsound, since (T) has some untrue instances. On the other hand, it is characteristic of paraconsistent solutions to diagnose sorites arguments as sound but not valid, since every instance of (T) is true but the rule of *modus ponens* is not valid. However, the weak paracompleteness/paraconsistency of supervaluationism and subvaluationism makes them agree where K3 and LP disagree. According to K3 the tolerance principle is untrue; according to LP is true; however, according to both supervaluationism and subvaluationism the tolerance principle is superfalse (note that this reflects the subvaluationist failure of adjunction and, more generally, universal generalization).[7]

Mixed consequence such as *st* and *pb* share features with both paracomplete and paraconsistent approaches. For both *st* and *pb* the argument in its B-form is unsound. For both, every instance of the tolerance principle is valid. The difference in weak/strong paracompleteness and paraconsistency reveals in that although the tolerance principle is *st*-valid it is not *pb*-valid. Thus, although the solution to the paradox is of a similar kind, the situation is somewhat different. The solution to B-type arguments according to *st* involves the claim that a valid sentence cannot always be used as a premise for a valid argument. Since although (T) is *st*-valid (since it is *t*-valid) it cannot be used in as a premise is an *st*-valid argument ((T) would need to be *s*-valid, which is not). The solution to this formulation of the paradox according to *pb* avoids this problem, but at the price of admitting the subvaluationist characteristic failure of universal generalization.

[7]See (Dietz, 2010, sect. 5) for a lucid presentation of the solution to the paradox concerning K3, LP, supervaluationism and subvaluationism.

Conclusion

In this paper we have characterized the space of consequence relations that we might define out of the three notions of satisfaction: local truth, supertruth and subtruth. For the restricted vocabulary, mixed notions of consequence that go from higher to lower standards of satisfaction (*pb*, *lb* and *pl*) coincide with local consequence. The remaining relations are all weaker than local consequence with *bp* as the weakest possible relation in the present setting (this consequence relation holds only if either the premises contain a classical contradiction or the conclusions a classical validity). We turned to see how similarity relations can be connected to the present framework. A natural idea is that similarity guarantees at most a small difference in semantic status. Thus, for example, the local truth of Pa in an ST-model guarantees the subtruth of Pb for any similar b in the model. When we allow similarity predicates into the language, all the consequence relations are distinct. Among the three stronger notions of logical consequence, *pb* seems to be the best option since *modus ponens* is valid and the deduction theorem holds. Though, due to their classicality, none of the logics discussed in this paper validate the tolerance principle, we can nonetheless provide a tolerant solution to the sorites paradox for its step-by-step formulation.

In the last section we briefly compared the notions of logical consequence *st* and *pb*. Both notions provide a satisfactory solution to the formulation of the sorites involving similarity relations and a chain of arguments. However, they differ on the solution to the formulation involving the tolerance principle. Both agree on the idea that the argument is unsound, and both agree that each instance of (T) is valid but while for *st* (T) is valid, for *pb* is not. The first is committed to the idea that a valid sentence might not qualify as a good premise for a valid argument; the second avoids this consequence at the price of endorsing the subvaluationist characteristic failure of universal generalization.

It looks to us that the mentioned differences between *st* and *pb* do not constitute enough evidence to tilt the balance towards any of them.[8] We already argued that *st* fits well with psycholinguistic data from recent experiments (Ripley (2009), Alxatib and Pelletier (201x) and Serchuk et al. (201x)) but we have not yet investigated whether the present framework fares well with these results. Other questions might be relevant to decide this issue. First, whether we can naturally introduce a *tolerant conditional* to provide a sound-but-unvalid solution to the formulation of the paradox involving (T) while preserving at the same time the properties of a good conditional such as *modus ponens* and the deduction theorem. Second, given

[8]See, however, Ripley (201x) for arguments against super- and subvaluationist non truth-functionality.

a natural formulation of the notion of a borderline case, whether we can accommodate the phenomenon of higher-order vagueness (an issue we also left for future work in TCS).

Appendix: tableaux

We can provide a tableaux system to check for any of the consequence relations presented in this paper. The idea is based on the obvious analogy of S-models and models of first-order modal logic with constant domain and a universal accessibility relation.

Definition 4 (Global modality)**.** *For any S-model $\langle \mathscr{M}, M\rangle$: $\mathscr{M}, M \vDash^l \Box\varphi$ iff $\forall M \in \mathscr{M},\ M \vDash \varphi$. $\mathscr{M}, M \vDash^l \Diamond\varphi$ iff $\mathscr{M}, M \vDash^l \neg\Box\neg\varphi$.*

Lemma 8. *For any S-model $\langle \mathscr{M}, M\rangle$: $\mathscr{M}, M \vDash^l \Box\varphi$ iff $\mathscr{M}, M \vDash^p \varphi$ and $\mathscr{M}, M \vDash^l \Diamond\varphi$ iff $\mathscr{M}, M \vDash^b \varphi$.*

Proof. From the definitions. □

These connections give us a neat way to apply standard modal tableaux for any of our nine notions of consequence.[9] Suppose we want to check whether $\Gamma \vDash^{pb} \Delta$. Then we have to construct a standard tableau for $\Box(\Gamma) \cup \neg\Diamond(\Delta)$. In our adaptation of modal tableaux, the nodes of a tree are something of the form φ, i where φ is a formula and i is a natural number (numbers designate classical models in an ST-model). The rules corresponding to $\Box$ and $\Diamond$ are:

$\Box\varphi, i$	$\Diamond\varphi, i$
φ, j (for any j in the tableau)	φ, j (for a new j)

We should further consider particular rules for similarity predicates. Given the characterization of these expressions, these are the corresponding rules:

$Pu, 0$ uI_Pv	$\neg Pu, 0$ uI_Pv
Pv, i (for a new i)	$\neg Pv, i$ (for a new i)

[9]See Priest (2008) for modal tableaux.

Here 0 is is our "designated model". Similarity claims are always interpreted in a fixed way, that is why there is no need of tagging the corresponding lines. Accordingly, Boxes and Diamonds should have no effect over similarity claims. Finally, we consider a rule for the symmetry of $\sim_P$ (we do not need to introduce a rule for the reflexivity of $\sim_P$ since in any tableau the node $Pa, 0$ will always lead to a stronger claim than the claim according to which Pa holds at some accessible m, namely, to the claim that Pa holds at accessible 0):

$$\boxed{\begin{array}{c} uI_Pv \\ \downarrow \\ vI_Pu \end{array}}$$

Example 1 $aI_Pb \vDash^{pb} Pa \rightarrow Pb$

$$\begin{array}{c} aI_Pb \\ \neg\Diamond(Pa \rightarrow Pb), 0 \\ \Box\neg(Pa \rightarrow Pb), 0 \\ \neg(Pa \rightarrow Pb), 0 \\ Pa, 0 \\ \neg Pb, 0 \\ Pb, 1 \\ \neg(Pa \rightarrow Pb), 1 \\ \neg Pb, 1 \\ \otimes \end{array}$$

References

Alxatib, S. and Pelletier, J. (201x). The psychology of vagueness: Borderline cases and contradictions. *Mind and Language*, Forthcoming.

Arruda, A. (1989). Aspects of the historical development of paraconsistent logic. In Priest, Routley and Norman (eds.) *Paraconsistent logic: essays on the inconsistent.* Munchen: Philosophia.

Cobreros, P., Egre, P., Ripley, D., and van Rooij, R. (2010). Tolerant, classical, strict. *Journal of Philosophical Logic*, Forthcoming.

Dietz, R. (2010). Vagueness and indeterminacy. In L. Horsten and R. Pettigrew (ed.) *The Continuum Companion to Philosophical Logic.* Oxford University Press., forthcoming.

Fine, K. (1975). Vagueness, truth and logic. *Synthese*, 30:265–300.

Hyde, D. (2008). *Vagueness, Logic and Ontology.* Aldershot: Ashgate.

Kremer, P. and Kremer, M. (2003). Some supervaluation-based consequence relations. *Journal of Philosophical Logic.*

Priest, G. (2008). *An Introduction to Non-Classical Logic: From If to Is.* Cambridge University Press.

Ripley, D. (2009). Contradictions at the borders. *Manuscript.*

Ripley, D. (201x). Sorting out the sorites. In Berto, F., Mares, E., and Tanaka, K. (eds.) *Paraconsistent Logic (tentative title)*, Forthcoming.

Serchuk, P., Hargreaves, I., and Zach, R. (201x). Vagueness, logic and use: Four experimental studies on vagueness. *Mind and Language*, Forthcoming.

van Rooij, R. (2010). Vagueness, tolerance and non-transitive entailment. *Manuscript.*