J Log Lang Inf (2013) 22:391–420
DOI 10.1007/s10849-013-9183-7

Vagueness and Order Effects in Color Categorization

Paul Egré · Vincent de Gardelle · David Ripley

Published online: 13 December 2013

Abstract This paper proposes an experimental investigation of the use of vague predicates in dynamic sorites. We present the results of two studies in which subjects had to categorize colored squares at the borderline between two color categories (Green vs. Blue, Yellow vs. Orange). Our main aim was to probe for hysteresis in the ordered transitions between the respective colors, namely for the longer persistence of the initial category. Our main finding is a reverse phenomenon of *enhanced contrast* (i.e. negative hysteresis), present in two different tasks, a comparative task involving two color names, and a yes/no task involving a single color name, but not found in a corresponding color matching task. We propose an optimality-theoretic explanation of this effect in terms of the strict-tolerant framework of Cobreros et al. (J Philos Log 1–39, 2012), in which borderline cases are characterized in a dual manner in terms of overlap between tolerant extensions, and underlap between strict extensions.

P. Egré (✉)
Institut Jean-Nicod (CNRS-ENS-EHESS), 29, rue d'Ulm, 75005 Paris, France
e-mail: paulegre@gmail.com

P. Egré
NYU, Department of Philosophy, 5 Washington Place, New York, NY 10003, USA

V. de Gardelle
Laboratoire de Psychologie de la Perception, Universite Paris Descartes, CNRS,
45 rue des Saints-Peres, 75006 Paris, France
e-mail: vincent.gardelle@gmail.com

D. Ripley
University of Connecticut, Department of Philosophy,
344 Mansfield Road, Storrs, CT 06269-1054, USA
e-mail: davewripley@gmail.com

Keywords Vagueness · Borderline cases · Dynamic sorites · Color categorization · Hysteresis · Enhanced contrast · Order effects · Strict-tolerant semantics · Contextualism · Contradictions

1 Introduction

Vague predicates of natural language such as “tall”, “rich” or “red” are often characterized by two related features: they have borderline cases of application, and they are sorites-susceptible. To say that they have borderline cases of application is to say that there are cases for which we feel hesitant between applying and denying the predicate of the object. To say that they are sorites-susceptible means that such predicates seem relatively insensitive to small differences: if a predicate like “tall” applies to an individual, then it seems equally applicable to an individual whose height differs very little.

While borderlineness and sorites-susceptibility are widely accepted as two main symptoms of vagueness, a lot of disagreement persists about the semantic status of borderline cases on the one hand, and about the correct analysis to give of the phenomenon of sorites-susceptibility on the other. Regarding borderline cases in particular, at least three broad families of views can be opposed, which we may call: the *sharp boundary view*, the *underlap view*, and the *overlap view*. The sharp boundary view is the view on which vague predicates have precise underlying boundaries, but about whose location we are uncertain. Borderline cases on that view have no special semantic status, they only have a special epistemic status (see in particular Williamson 1994). The underlap view is the view on which vague predicates are semantically indeterminate, leaving a gap between the extension of the predicate and the extension of its negation: borderline cases are those that fall in the gap (see for example Fine 1975). The overlap view finally is the view on which vague predicates correspond to categories that can encroach on each other (Hyde 1997), or even fade into one another (viz. Smith 2008). Borderline cases on that view are ‘glutty’ cases, lying at the intersection of distinct categories.

Several experimental studies have been conducted to adjudicate between these different views of borderline cases (see Bonini et al. 1999; Ripley 2011; Alxatib and Pelletier 2011; Serchuk et al. 2011), with new evidence in recent years in reappraisal of the overlap view of borderline cases (a view initially demoted in Bonini et al. 1999). Ripley in particular found that faced with a set of circle–square pairs at various distances from one another, participants readily report “the circle is and isn’t near the square” for the middling pair. Likewise, Alxatib and Pelletier find that faced with a set of individuals of various heights, participants readily report “he is tall and not tall” for the man of middling height. And Serchuk et al. establish that a significant proportion of participants asked to imagine a woman “between the clearly rich and the clearly not rich people” report that she is “rich and not rich”.

Although such studies give us information about the semantic representation of borderline cases for contexts of static comparison, they do not contain information about the phenomenon of sorites-susceptibility proper, that is about the way in which a semantic judgment concerning one case can dynamically affect the judgment concerning a similar case. Such conditions, in which subjects have to issue step by step judgments along a sequence of gradually differing objects, are commonly referred to as 'forced marched sorites' or 'dynamic sorites' in the philosophical literature (see Horgan 1994; Raffman 1994, 2005).[1] In an unpublished pilot study, Raffman and colleagues investigated both phenomena jointly (cf. Lindsey et al. 2005, cited in Raffman 2005, 2009). In their experiment, participants were presented with a series of gradually shifting shades interpolated between a clear blue and a clear green (see Figs. 1, 2 below for comparable stimuli in our present study). Individual shades were shown consecutively in various orders, including random order, ascending from green to blue, and descending from blue to green, with the choice to categorize each shade as "BLUE", "GREEN", or to opt out with "?". In two additional ordered conditions (called "reversal" conditions), when going from blue to green (or from green to blue), as soon as the subjects switched semantic category (viz. from BLUE to GREEN), they were shown the immediately preceding shades in reverse order of appearance. What was found is that subjects, immediately after switching from BLUE to GREEN, continued to apply GREEN to the shades categorized as BLUE just before that, only switching back to BLUE later.

Raffman's data too may therefore be seen as evidence for an overlap view of vague predicates, since there is a range of shades that subjects equally call BLUE and GREEN depending on the context of the transition. More specifically, Raffman presents the data as exemplifying a phenomenon of local hysteresis. Hysteresis happens when, in an ordered condition, subjects switch category at a later position depending on the category they come from (Hock et al. 1993; Kelso 1995). For instance, in an ordered series of shades from blue to green, there is hysteresis if the switch happens at a greener shade when going from blue to green than in a random presentation order, and symmetrically if the switch happens at a bluer shade when going from green to blue. Hysteresis is a phenomenon that can be observed in various categorization processes, and in particular in the perception of ambiguous or multistable stimuli (see Hock et al. 1993, 2004). It can be defined, more abstractly, as the longer persistence of the category one is coming from in the transition toward a distinct category. Not all transitions between two stable categories manifest hysteresis, however. In some cases, subjects show a *critical boundary*, that is they could reliably switch at about the same position, irrespective of the order in which stimuli are presented. In other cases, subjects show *enhanced contrast* (also called *negative hysteresis*, *counterhysteresis* or *reverse hysteresis*), that is, they switch at an earlier position rather than at a later position in the ordered conditions (cf. Tuller et al. 1994; Kelso 1995) for a comparison between all three phenomena in the perceptual domain).

In this paper we propose to investigate this phenomenon of category switch in the dynamic transition between two vague categories, in order to shed further light on the semantic status of borderline cases. Our first goal is to test whether hysteresis is a robust phenomenon for the kind of task used by Raffman and colleagues. For one thing, Raffman et al.'s pilot study does not present a standard measure of hysteresis for the basic ordered conditions, but only indirect evidence based on the behavior of a few

[1] The expression 'dynamic sorites' is used to make a distinction with the 'regular sorites argument', viz. (Smith 2008), namely the abstract argument whereby a contradiction is drawn from the assumption that a soritical series is given, with the first individual being clearly P, the last being clearly not P, and one grants the tolerance principle that x_{n+1} is P in the series whenever x_n is P.

subjects in the specific reversal conditions. For another, even assuming hysteresis to be a robust phenomenon, their task does not allow us to establish whether such hysteresis is primarily driven by perceptual processes, or whether it is specifically linguistic and tied to the semantic content of the categories under comparison. In order to control for this, our first Experiment (Experiment 1) involves two tasks: a perceptual task in which subjects have to compare intermediate color samples between two categories to two templates taken to be clear instances of the categories in question, and a linguistic task in which subjects are asked to predicate the names of the color categories of the same color samples.

A second and related goal of our research is to probe for the adequacy of the semantic characterization of borderline cases in terms of an overlap between categories, rather than in terms of underlap or even in terms of sharp underlying boundaries. In the results of Alxatib and Pelletier, for example, but also in that of Serchuk et al., borderline cases are categorized not exclusively as “P and not P”, but also as cases that are “neither P nor not P”. This suggests that some notion of underlap is also in play in the way borderline cases are conceived. Thus, in the semantic account of vague predicates proposed in Cobreros et al. (2012), which we use as our framework in what follows, borderline cases are characterized more abstractly in a dual manner, as cases that are “P and not P” in one sense (the *tolerant* sense of P), but also as cases that are “neither P nor not P” in another sense (the *strict* sense of P). The question for us is whether any evidence for this notion of underlap can be expected to show up in a dynamic setting in terms of a specific order effect. Kalmus (1979) presents a task of color naming in which subjects evidenced a phenomenon of enhanced contrast, rather than hysteresis: that is, in the gradual transition between two colors, subjects switched category closer to the category they were coming from, rather than later.[2] As pointed out by Kalmus, however, and contrary to what happens in Raffman’s design, subjects were informed in advance of the direction of the transition, and the delay between consecutive shades was quite large, leaving room for a phenomenon of ‘forward-looking’. The expectation at the start of our research is therefore the opposite: that if Raffman’s data is robust, we should see hysteresis, rather than enhanced contrast, working with shorter delays between consecutive presentations than Kalmus, and not explicitly instructing our subjects about the direction of transition.

Our third goal in this research concerns the precise articulation of the notion of overlap between categories. The first question to ask is: *overlap between what?* Two notions of overlap need to be distinguished. One is the notion of overlap between distinct concepts (viz. Blue vs. Green). The second is the notion of overlap between a concept and its negation (viz. Blue vs. Not Blue). Our study compares the two notions. In Experiment 1, we essentially replicate Raffman’s design, by asking our subjects to select between two distinct colors names, thereby considering only the first notion of overlap. In Experiment 2, we instead asked our subjects to give a Yes/No answer to the linguistic description of the same color sample. In the same Experiment, furthermore, we ask our subjects to give Yes/No answers for two kinds of conjunctive descriptions of the same sample, namely “Blue and Green” and “Blue and not Blue”. A second

[2] We are indebted to Diana Raffman and to the referee of a previous paper for bringing Kalmus’ paper to our notice (see Egré 2011).

question we propose to clarify is: *overlap, or contextual shift between categories*? One way of understanding the notion of overlap is indeed purely in terms of a contextual shift between two mutually and exhaustive categories coming with a sharp boundary. On that view, each sharp category gets dragged into the other depending purely on the direction of the transition.[3] Another possibility is that the overlap is more than a phenomenon of boundary displacement, but that we first recognize a proper overlap area between two categories, and that this overlap region itself gets to be displaced depending on the direction of change. If that possibility is right, we may need to distinguish two effects: (static) overlap on the one hand, and (dynamic) contextual shift on the other. To select between these two hypotheses, our study investigates whether order effects can occur also for conjunctive predicates of the form "Blue and Green" and "Blue and not Blue".

2 Experiment 1

2.1 Methods

2.1.1 Participants

19 participants were recruited for our experiment, reporting normal or corrected vision, and no color-blindness, all of them being native speakers of French or fluent in French. In Experiment 1, as well as in Experiment 2 below, all instructions were given in French. The experiment took place at the Laboratoire de Sciences Cognitives et de Psycholinguistique in Paris (LSCP, DEC-ENS). Participants were paid for their participation (10 euros for one hour; the experiment lasted about 50mns on average).

2.1.2 Stimuli

We defined two sets of colors, one consisting of yellow to orange shades, the other of blue to green shades. Each color set consisted of 15 distinct color shades (see Figs. 1, 2). RGB values for the respective end shades of the Yellow–Orange set and Green–Blue set were defined between 0 and 1 as follows: Yellow = [1, 0.8, 0], Orange = [1, 0.4, 0], Blue = [0.1, 0.22, 0.78]; Green = [0.1, 0.78, 0.22]. The RGB values for the 13 intermediate shades in each set were defined by means of a regular and linear interpolation within each of the three RGB dimensions.

[3] See Raffman (2005) and the description given by (Smith, 2008, 118) of the response a contextualist might make to the finding of hysteresis:

> "Suppose we are walked along our Sorites series for F, and asked of each object in the series whether it is F, and then walked back the other way, and asked the same question of each object again. It is very likely that the point at which we stopped saying 'Yes' on the way out would be further along the series than the point at which we started saying 'Yes' on the way back. This behavior might seem rather difficult to explain on the recursive truth gap view, but it is easily explained by the contextualist: as we classify an object as F, we thereby make it true that that object and all others very similar to it—including the next object in the series—are F. We thus push the boundary between the F's and the borderline cases out before us as we go—and on the way back, we push the boundary back the other way".

Fig. 1 The Yellow–Orange color set

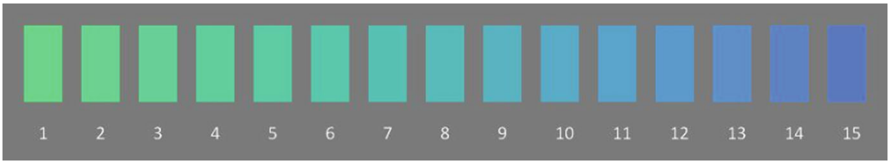

Fig. 2 The Green–Blue color set

The interpolation ensured equidistance between shades in the RGB encoding, but not necessarily equidistance in the participants' perceptual space. However, the choice of endpoints for each continuum was made so as to approximately get as many Yellow shades as Orange shades, and as many Blue shades as Green shades, based on this interpolation, and so as to get a smooth color transition between the two end shades in each set. The choice of the two color sets was made based on the comparison between several alternative color sets obtained by the same method, and relying on the consistency between the experimenters' evaluations on the test screen. We did not attempt to achieve congruent perceptual distances between adjacent shades across the two color sets (for instance we did not try to ensure that the distance between stimulus 10 and 11 be felt as exactly congruent for the Blue–Green and for the Yellow–Orange sets). Rather, our main reason for using two color sets was to see if we would find the same order effects replicated over and above such differences. A further reason was to introduce a variety of colors and to alternate between the two sets in order to reduce memory effects attached to each set.

Except in the posttest, where participants saw Figs. 1 and 2 above, shades were presented as squares of width = 100 pixels, either as one top color square versus two bottom color squares, or as a single top color square versus two bottom color names. The topmost square was always presented at 1/2 of the width of the screen, and 1/3 of the height. The bottom squares or bottom color names were always presented at 1/3 and 2/3 of the width of the screen respectively, and at 2/3 of its height.

All tests were run using Matlab's Psychtoolbox, in one test booth, on a 17" CRT screen refreshed at 70 Hz, and viewed from a 60 cm distance. Participants were not placed in total obscurity but in a lightened environment with fluorescent bulbs.[4] All stimuli were presented on a grey background with RGB coordinates [0.2, 0.2, 0.2]. participants used an AZERTY keyboard to give their answers. In Experiment 1, we

[4] In a first pilot study, all participants were put in a dark environment over a black background: however, participants reported more visual fatigue, and this had the undesirable effect of creating too much contrast between the color and its background.

Fig. 3 An instance of the perceptual task for the Yellow–Orange set

used a constant time lag between each answer and the next stimulus, of 200 ms, with a time lag of 500 ms between two consecutive blocks.

2.1.3 Task

The experiment consisted of three parts, which we refer to as the 'perceptual' task, the 'linguistic' task and the posttest.

- In the 'perceptual' task, participants were presented with three color squares, a top square and two squares below it, all three of them taken from the same color set (Blue–Green or Yellow–Orange). The two bottom squares always stood as the opposite endpoints of each color set, with their relative position (left or right) counterbalanced across trials (each square appearing left or right half of the time, but at random). The top square however could be any of the 15 squares in the corresponding color set (see Fig. 3). The instruction given to participants was to select, relative to each top square, the square below it to which they judged the top square to be more similar.
- In the 'linguistic' task, participants were presented with a single color square and two color names. The color names were either 'JAUNE' (Yellow) versus 'ORANGE' (Orange) for shades belonging to the Yellow–Orange spectrum, or 'BLEU' (Blue) versus 'VERT' (Green) for shades in the Blue–Green spectrum (see Fig. 4), with the relative position of the two color names counterbalanced across trials. The instruction given to participants was to select, relative to each presented color square, the color name that they judged to provide the best description of that color.

  In the 'linguistic' as well as in the 'perceptual' task, participants made their selection with the keyboard by pressing 'S' for left or 'L' for right.

- In the posttest, participants were shown the two color sets consecutively. For the first time, they were shown the 15 color shades from each set together on the screen, ordered from Green to Blue or from Blue to Green, and from Yellow to Orange or from Orange to Yellow, depending on the participant, as in Figs. 1 and 2 above. For each color set, they were asked four questions for which they had to provide their answer on a separate sheet of paper. The questions, for each color

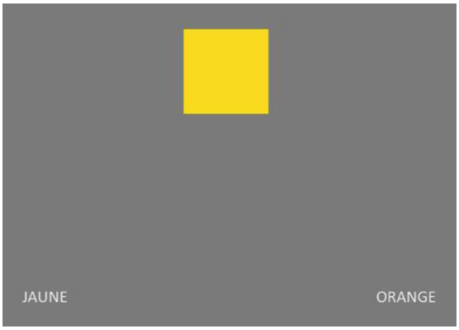

Fig. 4 An instance of the linguistic task for the Yellow–Orange set

set, were: 1) *Which shade seems to you to be the Bluest?* 2) *Which shade seems to you to be the Greenest?* 3) *Where would you locate the boundary between Blue and Green?* 4) *Which shades seem to you to be between Blue and Green?*, and similarly for Yellow versus Orange. Finally, participants were invited to give free comments on the experiment at the back of the sheet.

2.1.4 Design

We used a 2x2x3 factorial design for our experiment: there were 2 tasks (‘perceptual’ vs. ‘linguistic’), 2 color sets for each task (B–G vs. Y–O) and 3 possible orders for the stimuli in each block (Random, Ascending, Descending). There were 15 stimuli per block, and 8 repetitions for each block, which made a total of 1440 trials for each participant.

Participants started either with the perceptual or with the linguistic task, depending on the number they were assigned as participants. They moved to the next task only after they had completed the first, so the two tasks were not intertwined.

Each task was divided into blocks of 15 trials, one for each stimulus in each color set. There were 3 possible orders in each block: participants were shown the 15 shades in each color set consecutively, either in random order, or in ascending order (from Blue to Green, from Yellow to Orange), or in descending order (from Green to Blue, from Orange to Yellow).

In the perceptual task, for example, an “ascending” block was an ordered series of 15 presentations of a color triad, consisting of the top square gradually going from Yellow to Orange (or from Blue to Green), with the bottom squares always as the Yellowest and most Orange squares in the spectrum (respectively the Bluest and Greenest). In the linguistic task, similarly, a “descending” block was an ordered series of 15 single color presentations, gradually going from Orange to Yellow (or from Green to Blue), each time with the same color names to select from, that is JAUNE versus ORANGE (respectively BLUE vs. VERT).

Finally, each block in one color set was followed by a block in the other color set, so as to minimize memory effects and other side effects of order when going from one block to the next.

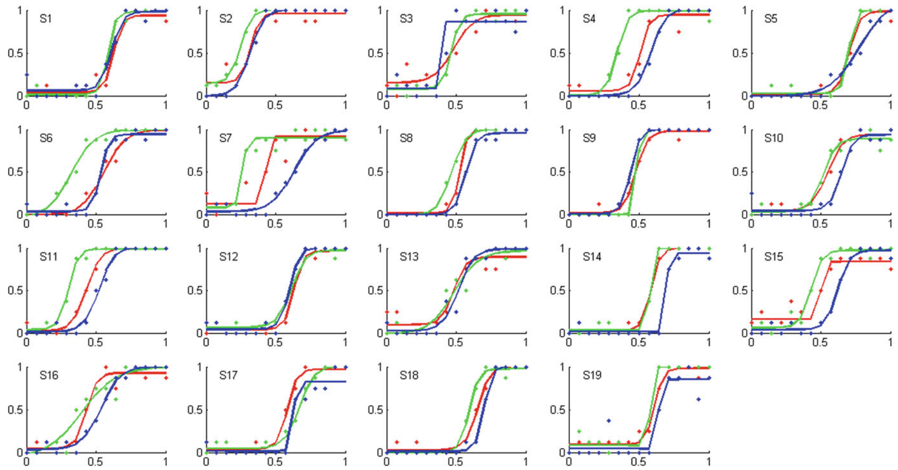

Fig. 5 Raw data of the linguistic task for the Yellow–Orange color set (*fitted curves*); In *Red*: random order; *Green*: ascending; *Blue*: descending

2.2 Results

Our main variable of interest was the position of inflection points in the psychometric curves. In particular, hysteresis effects would be reflected by changes in the inflection points as a function of the order of presentation of the stimuli.

For each subject and for each condition of our design (order × task × stimulus set), we computed a psychometric curve as the proportion (over trials) of orange (respectively blue) responses, as a function of the position on the color spectrum. These curves were individually fit by a sigmoid function with 4 parameters as per the following equation, where x_0 is the inflection point, a and b are the asymptotic limits, and s is half the inverse of the slope at the inflection point:

$$\psi(x) = a + \frac{b}{1 + e^{\frac{-(x-x_0)}{s}}}$$

An example of the resulting psychometric curves is given in Fig. 5 for the linguistic task in the Yellow–Orange set. As a preliminary inspection reveals on that figure, the curves of a few subjects show a pattern typical of negative hysteresis (viz. S7 or S11). In order to test for order effects, we first subjected our (fitted) inflection points to an analysis of variance (ANOVA) with order, task, stimulus set as within-subject factors, and subject as a random factor. The data included in this analysis is represented on Fig. 6. The ANOVA revealed main effects of order ($F(2, 19) = 8.87$, $p < 0.001$), stimulus set ($F(1, 19) = 44.73$, $p < 0.001$), and interaction between task and order factors ($F(2, 19) = 33.09$, $p < 0.001$). No other effects or interactions between factors were found.

The interaction between order and task reflects a negative hysteresis effect that is present only in the linguistic task. Indeed, when we then computed separate ANOVAs

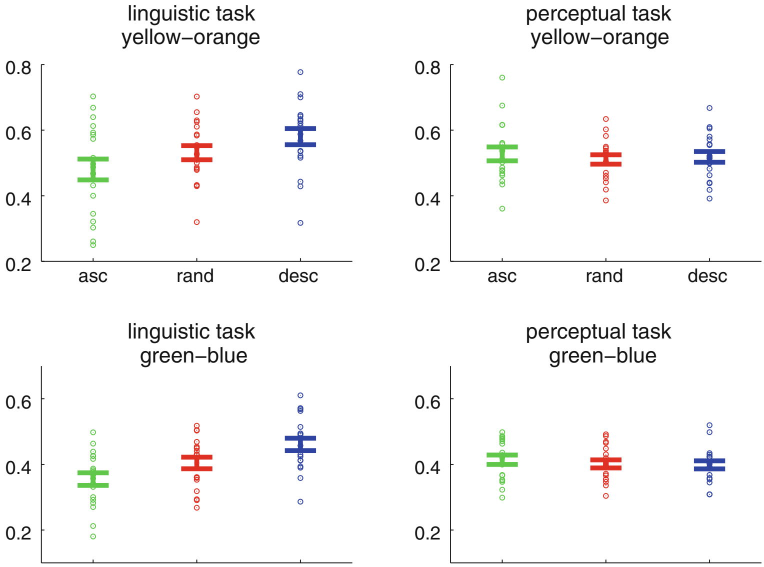

Fig. 6 Mean position of the inflection points. In *Red*: random order; *Green*: ascending; *Blue*: descending

for each task, with order and stimulus set as within-subject factors, we found no effect of order ($p > 0.3$) in the perceptual task, but an effect of order in the linguistic task ($F(2, 19) = 24.78$, $p < 0.001$). Additionally, these analyses revealed main effects of stimulus set in both tasks (perceptual: $F(1, 19) = 46.11$, $p < 0.001$; linguistic: $F(1, 19) = 27.06$, $p < 0.001$), but no interactions between stimuli and order ($p > 0.7$ in both tasks).

To complete these analyses, we assessed the effect of order in the two stimulus sets in the linguistic task, and found that it was significant in both the Yellow–Orange stimulus ($F(2, 19) = 12.37$, $p < 0.001$), and the Green–Blue stimulus set ($F(2, 19) = 30.55$). This confirms that the negative hysteresis effect found in the linguistic task was present for both color sets.

2.3 Discussion

In this experiment we compared the subjects sensitivity to the order of presentation of color stimuli in two distinct tasks of color categorization. In the first task, participants were instructed to best match a color stimulus to one of two stimuli, each representing one end of the spectrum from which the sample stimulus was taken. In the second task, participants were instructed to best match a color stimulus to one of two color names, each naming one end of the spectrum relevant in both tasks.

Our main finding in this experiment is a phenomenon of negative hysteresis in the linguistic task, meaning that participants shift linguistic category at an earlier position when they do a gradual transition from one end of the color set to other. In the perceptual task, no significant order effect was found, but rather, a phenomenon of critical boundary, namely switch of categories at a relatively stable position in all conditions. These findings raise three main questions:

(1) how is it that a negative hysteresis was found in the linguistic task, rather than a standard hysteresis?
(2) How is it that no order effect was found in the perceptual task?
(3) Where does the contrast originate between the linguistic and the perceptual task?

As explained in introduction, our expectations for this experiment were opposite. First of all, if anywhere, we expected to observe hysteresis in the perceptual task. This expectation was based in large part on the supposition that intermediate color shades might pattern as ambiguous stimuli, hysteresis being generally considered as the signature of multistable perception (see Kelso 1995; Egré 2009, 2011). Secondly, our basic expectation was to see parallel effects in the perceptual and the linguistic task, based on the structural analogy of the tasks and on Raffman's own results.

Whether the lack of order effect in the perceptual task is a robust finding (question 2) is not something our experiment can tell. In particular, we cannot rule out the possibility that a more fine-grained transition between the end shades of our stimuli (using, for instance, 20 shades instead of 15) might have an effect on the perceptual dynamics, giving rise to positive hysteresis. On the other hand, we see the lack of this order effect in each of our two stimulus sets. Supposing such a lack of order effect to be robust, one possible explanation for it is that upon each trial, because participants can see the opposite ends of the spectrum in front of them, they are able to estimate the boundary between those in a reliable way, and in particular they are able to issue reliable judgments of relative similarity.[5]

The same does not hold of the linguistic task. In the linguistic task, subjects do not have matching stimuli perceptually available on each trial to which the sample stimulus can be externally compared. Instead, they have to rely on their memory of colors attached to the names "Vert", "Bleu", "Jaune" and "Orange" to decide whether it is more adequate to call the color sample "Vert" rather than "Bleu", or "Jaune" rather than "Orange". In other words, a possible explanation for the contrast in behavior between the perceptual and the linguistic tasks (question 3) concerns the fact that names of colors do not present concrete colors, but require participants to rely on their semantic memory of perceived colors.

What remains to be explained, however, is why a negative hysteresis was found in the linguistic task, rather than a positive hysteresis (question 1), or even, rather than no order effect at all. The first thing to point out about this result is that it is

[5] (Kamp 1981, 241) made a thought experiment in which he envisaged two situations of forced march from a clear green to a clear yellow: one in which all stimuli, including the ends of the color set, are always in view, and one in which color shades are only shown one by one. He predicted (with caution, see his fn. 7) that in the second situation, subjects might "carry on answering "green" for a longer time" than in the first. We do not have here a proper counterpart of his first scenario to really check the prediction. In the light of our data, this would imply that the enhanced contrast should be even more pronounced in the first setting than in the second, but this is not what Kamp intended. Rather, Kamp predicted that the lack of anchor points in the second case should favor hysteretical behavior past the middling shade. Judging from our findings, we see that the effect of anchoring points may indeed be to secure a critical boundary, as Kamp imagines for his first experiment, but their absence would have the counterhysteretical effect of subjects restricting the extension of the category they come from.

consistent with the finding of Kalmus (1979), who also observed negative hysteresis in a task in which participants had to name a color sample along similar transitions by selecting a color name from a pre-defined set. Kalmus did not propose any general explanation for his effect, but attributes it to two potential confounds: the long duration between successive trials, and the participants' explicit awareness of the direction of the transition. In our experiment, durations between trials were much shorter than in Kalmus's study. In the first ascending and descending blocks, on the other hand, while our participants may not have been aware directly of the order of the transition, debriefing makes it clear that after a few repetitions, they were able to realize when they were in an ordered condition. But even so, the awareness of the direction of the transition by itself is not sufficient to explain why participants would shift category earlier rather than later, since the same awareness was presumably available to our participants in the perceptual task.

Various explanations for this effect are possible. One line of explanation would be to assume that "Orange" and "Yellow" semantically and psychologically exclude each other. On an epistemic version of this view, subjects would name color samples based on the uncertain evaluation they make of the position of a sharp boundary between "Orange" and "Yellow". For instance, when starting a transition from Orange to Yellow, subjects would first feel quite confident that the first few shades are "Orange". When confronted with an intermediate shade, they would start feeling uncertain of which side of the boundary it lies on Williamson (1994), Bonini et al. (1999). Consistently with an assumption made by Bonini et al. (1999), they may decide that it is preferable to no longer commit to the category for which they have issued reliable judgments. For lack of a better choice, they would switch to the other category, as soon as they experience uncertainty regarding the shade. Alternatively, on a supervaluationist articulation of the exclusion hypothesis, subjects would first use "Orange" for the determinate cases of Orange they see. When hitting upon the first noticeably borderline case, subjects would no longer be in a position to use "Orange". Again, they may decide that it is preferable to (incorrectly) call a borderline case "Yellow" then, rather than to distort the correct use of "Orange" they were making thus far.

Both explanations are plausible concerning the idea that the transition from clear to unclear cases of a category plays a role in switching color judgments, but the hypothesis of either a sharp line or of an exclusion region between color categories strikes us as less plausible than the assumption of an overlap (for reasons we will only fully explain in the next section). One way of articulating this overlap view is the one explored in Cobreros et al. (2012). According to that theory, a vague predicate can be used in two ways, a "strict" and a "tolerant" way. In its strict use, a predicate is meant to apply to clear or unambiguous instances of the category. In the case of a color predicate like "Jaune", for example, "Jaune" is used strictly only to denote a good or prototypical instance of yellow. In its tolerant use, by contrast, a predicate can be used in a looser sense, including for cases that are not perfect instances of the predicate. For example, "Jaune" can be used tolerantly to denote a shade that has a sufficient degree of yellowness in it, even if it is not the maximum degree. Importantly, the theory admits that "Jaune" and "Orange" can tolerantly overlap, although they strictly underlap. Assuming a range of shades for which Yellow and Orange overlap, what is plausible in the ordered conditions is that subjects first anchor the use of the color predicate they start from to clear cases, hence use it in a strict sense. The negative hysteresis found in the linguistic task suggests that subjects prefer to use a new predicate, rather than to tolerantly reinterpret a predicate they have been using strictly first.

Assuming an overlap view of color categories, we can give a principled explanation of this pattern of behavior, based on essentially three principles, namely two pragmatic principles about information communication, and a third principle regarding the representation of colors proper. The two pragmatic principles are principles of *accuracy* and *informativeness*. The accuracy principle we are assuming may be viewed as a version of Grice's maxim of Quality, and is the principle that participants should report a shade as of a given color only if the shade is indeed of the color. The informativeness principle may be related to Grice's maxim of Quantity, and basically implies that subjects should try to be informative of a change of color quality when they detect such a change. The third assumption we are making finally concerns the representation of the color spectrum. This is the hypothesis that subjects apprehend each color continuum as divided into essentially three regions (with fuzzy contours), for example, in the case of the Yellow–Orange set, a region of clear Yellow, an overlap region between Yellow and Orange, and a region of clear Orange.

Combined together, those three principles can explain why it is rational to switch earlier rather than later in the description of the color. In an ordered sequence starting with a clear yellow square, subjects cannot select "Orange" for the first square, on pain of issuing a false judgment. Selecting "Jaune" is therefore the most informative and the only accurate choice, hence the use of "Jaune" is optimal over "Orange". When they reach the first few noticeably Orange–Yellowish shades, on the other hand, selecting "Jaune" again, though not inaccurate as understood tolerantly, would impart that the shades have the same yellow quality as the original shades, which appears less informative than to impart that they have a different quality. Accordingly, subjects should switch to the tolerant or loose sense of "Orange", which is then optimal. Finally, when entering a region of clear Orange, subject have no choice but to keep using "Orange", on pain of flouting the principle of accuracy (they obviously won't say "Jaune" for a clear Orange). They can't be as informative as they might otherwise be with more expressive resources at their disposal, but at least they are accurate.

This explanation no longer has to rest on the assumption that subjects are uncertain about the position of an underlying sharp boundary, and it also dispenses with the exclusion view of color categories more generally. Instead, it rests fundamentally on the idea that subjects represent to themselves a middle region of overlap between two color categories, and that subjects try to minimize violations of the two pragmatic principles at each point. A representation of the relative satisfaction of the two conditions of accuracy and informativeness relative to each color name and region is given in Table 1. The distribution of + and − (for whether a condition is satisfied or not) and the computation of the optimal alternatives relies on two further specific assumptions: the first is that whether a name is informative relative to a region depends on whether it can be used to signal a change of color quality, *relative to the immediately previous optimal choice*. This essentially is what allows us to capture the dynamics of categorization in this case: in particular, 'Jaune' is here counted as informative in the Orange region, despite being inaccurate, knowing that 'Orange' is optimal in the previous Yellow–Orange region; and both 'Jaune' and 'Orange' are informative in the first region, since

Table 1 Satisfaction of accuracy and informativeness relative to the color region and color name (transition from Yellow to Orange)

Color region	Yellow		Yellow–Orange		Orange	
Color name	▹ *Jaune*	*Orange*	*Jaune*	▹ *Orange*	*Jaune*	▹ *Orange*
Accuracy	+	−	+	+	−	+
Informativeness	+	+	−	+	+	−

+ Satisfaction, − violation, and ▹ the optimal alternative relative to the region

they signal some color quality. The second assumption is that accuracy is a constraint ranking above informativeness. For example, in the region of clear Orange, "Orange" is a better alternative than "Yellow", for although "Yellow" would be more informative of a change given that "Orange" was used previously, it violates the higher-ranked constraint of accuracy.[6]

Two remarks should be made about our proposed explanation. The first is that, in principle, subjects might also try to anticipate that it might be more informative to use "Orange" when they enter the region of clear Orange rather than reuse a label they have described earlier. Such subjects would anticipate that if they reserve "Orange" only for the clear Orange shades, they will be both more informative and more accurate when they enter that region (against being only accurate). But such idealized subjects should also realize that it will be less informative to use Yellow in the middle region than to switch category to signal that it is not the same Yellow as initially. In other words, such subjects would try to make their last choice strictly better than the previous choice (that is better in all dimensions), but then at the expense of making a weakly dominant choice earlier. Our explanation therefore assumes that subjects are reasoning optimally at each local point.

Secondly, other accounts beyond the strict-tolerant account of vague predicates we favor would be compatible with the optimality-theoretic explanation given for the enhanced contrast reported so far. What really matters for the explanation given here is the hypothesis of an overlap between the color predicates "Yellow" and "Orange". In particular, it would be enough for the OT-explanation we have given to suppose that "Orange" and "Yellow" are always used in one and the same sense, but that they overlap. On that picture, "Orange" and "Yellow" may very well have sharp classical extensions, but encroaching on each other. In other words, the overlap idea, while probably not in the spirit of standard epistemicism (and while clearly excluded in supervaluationism), remains compatible with classical logic strictly speaking. Such a picture would even be more parsimonious than one relying on something like a strong versus loose meaning for an expression.

To adjudicate between that view and one relying on the strict-tolerant distinction, what we need is to look at judgments concerning the negation of such color predicates. Indeed, the strict-tolerant distinction is fundamentally driven by the consideration that

[6] An alternative way of articulating informativeness would be to assume that informativeness requires accuracy; this would change some value assignments in the table, but make the same prediction of an enhanced constrast in the present setting.

a more revisionary notion of meaning is required provided a predicate and its negation can be said to overlap (see Cobreros et al. 2012). If there are such cases in which overlap means overlap between a predicate and its negation, then the corresponding extensions can no longer be classical. In the next Experiment, we therefore consider whether similar order effects occur in judgments involving a predicate and its negation.

3 Experiment 2

While plausible to us, the explanation proposed at the end of the previous section supposes that the negative hysteresis found in the linguistic task is a robust effect, and moreover that subjects do indeed recognize an overlap region in each color set. To get confirmation, we ran a second experiment, with a different group of subjects. Unlike Experiment 1, Experiment 2 involves two linguistic tasks. This time, however, instead of giving a forced choice between two color names, participants were asked to agree or disagree with various sentential descriptions of the shade in terms of a single color name.

The first aim of this experiment was thus to see whether we would replicate the same order effects found in Experiment 1 with YES/NO judgments as applying to sentential descriptions of the form: “the square is yellow”, “the square is not yellow”, with the latter involving the predicate’s negation. When facing the choice between “Orange” and “Yellow”, participants in Experiment 1 have no way to report that a shade is not orange or not yellow. However, they may very well consider some shades to be neither Orange nor Yellow. For those shades, participants in Experiment 2 have more expressive resources at their disposal, since they can agree to “the square is not yellow”.

A second aim of the experiment was to see how participants would make use of conjunctive descriptions of the shades of the form “the square is yellow and orange” / “the square is yellow and not yellow” in a dynamic setting. If participants recognize a proper overlap region between categories, the prediction is that they should assent to such descriptions significantly more within the overlap region. What remains to be seen, however, is whether this overlap region too is susceptible to order effects. The hypothesis under examination is that if the borderline area between yellow and orange (respectively yellow and not yellow) is represented as a proper region, then we should observe similar order effects in the application of conjunctive predicates intended to describe the region.

3.1 Methods

3.1.1 Participants

We recruited 26 participants for this experiment, reporting normal or corrected vision, with no color-blindness, all fluent in French. The experiment took place at the Laboratoire de Sciences Cognitives et de Psycholinguistique in Paris (LSCP, DEC-ENS). The experiment lasted about 60 minutes and subjects were paid 10 euros for their participation.

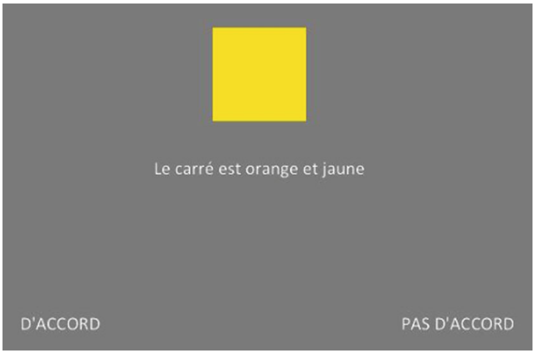

Fig. 7 An instance of a conjunctive sentence (Yellow–Orange set)

3.1.2 Stimuli

We used the same two color sets we had used in Experiment 1, with one set of colors ranging from yellow to orange, and another from blue to green, each consisting of 15 shades with the same physical characteristics. Participants were placed in the same environment as in Experiment 1.

3.1.3 Task

The experiment consisted of two parts, the main part and then the same posttest used for Experiment 1. The main part was divided in 96 blocks, of 15 trials each. In each trial, a single color square and a French sentence describing the square were presented (see Fig. 7), and subjects had to decide whether the sentence correctly described the square, and indicated their choice by selecting 'D'ACCORD' ("agree") or 'PAS D'ACCORD' ("disagree"). On each trial the choice options were randomly associated with key press 'S' and 'L' on the keyboard. In each block, the same sentence and the same color continuum was used in all 15 trials. Each block was separated from the next by a sentence asking to start a new sequence. Within each block, after each answer a random lag was introduced before the presentation of the next stimulus (mean duration 800ms, theoretially comprised between 550 ms and 1050 ms).

3.1.4 Design

We used a $2 \times 2 \times 2 \times 2 \times 3$ factorial design for our experiment, making a total of 48 conditions. First, we used two colour sets as before. Second, three factors manipulate the type of sentence used in one block: sentences could be *atomic* ('The square is P') or *conjunctive* ('The square is P and Q'), they could involve a negation or not ('the square isn't yellow', 'the square is yellow and not yellow' versus 'the square is yellow', 'the square is yellow and orange'), and they could involve one color label or the other for the same continuum (e.g. 'the square is yellow' versus 'the square is orange'). Finally, we used 3 order of presentation for the stimuli in each block: Random, from one end to the other end of the continuum, or in the reverse order. In atomic blocks, we call Ascending blocks those for which the color label used in the sentence should be considered as valid at the end of the block, and Descending the opposite blocks. In conjunctive blocks that rely on two color labels, we use the same convention, with respect to the first of the two labels. Each condition was used in two blocks of 15 trials, for a total of 96 blocks (1440 trials) per participant. All blocks were presented in a random order.

3.2 Results

One participant was excluded from analysis, having completed the experiment in much less time than the other participants, and exhibiting erratic responses.

3.2.1 Order Effects in the Atomic Sentences

Positive Atomic sentences First, we wanted to assess hysteresis effects in atomic and positive sentences to replicate our findings from Experiment 1. We thus computed psychometric curves by calculating the percentage of agreement with the predicate, for each participant, color set and order condition. To increase power, we collapsed across the two possible color labels while holding constant the convention that ascending refers to an acceptation of the color term at the end of the block. We then fit these empirical psychometric curves with a 4 parameter sigmoid (Eq. 1) to assess whether the inflection point (x_0) would be affected by the order of presentation of stimuli in a block. Some unrealistic values (e.g. inflection point outside the 0-1 range), probably arising from noise and sparse data were replaced by a default 0.5 (which was neutral with respect to the subsequent tests). Similar results were obtained when we removed the participants affected by this issue.

We carried out an analysis of variance (ANOVA) on the inflection points with order and color set as within-subject factors, and subject as a random factor. The main effect of order was significant ($F(2, 48) = 6.43$, $p < 0.005$), and there was no effect of stimulus set nor an interaction (both $p > 0.3$). The data included in this analysis is represented on Fig. 8. The effect of order indicated an anti-hysteresis effect, with inflection points being smaller and thus closer to the start of the block for ascending blocks (Yellow–Orange: 0.48 vs. 0.52; Green–Blue: 0.47 vs. 0.54) (Fig. 10). In other words, in ascending blocks participants started by disagreeing with the proposed color label but then made an early switch to agreeing, and in descending they started by agreeing and made an early switch to disagreeing. This pattern fully replicates the effects of Experiment 1.

Negative Atomic sentences Next, we assessed hysteresis effects in negative sentences, using the same approach (Fig. 9). Here, to compare responses across positive and negative sentences, we inverted the responses for negative sentences, thus treating agreement with “the square is blue” as equivalent to disagreement to “the square is green”. The ANOVA on the inflection points revealed main effects of order ($F(2, 48) = 3.40$, $p < 0.05$), with no main effect of stimulus set and no interaction (both $p > 0.4$). The effect of order indicated that in ascending blocks participants started by agreeing but made an early switch to disagreeing with the sentence, thereby

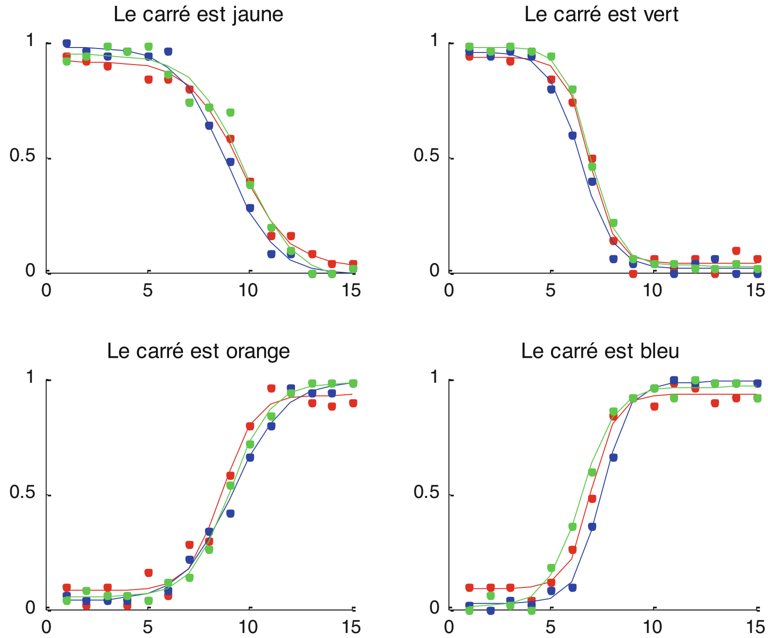

Fig. 8 Raw data for positive atomic sentences x-axis: position on the color continuum, from yellow to orange (*left plots*) or green to blue (*right plots*). y-axis: percentage of 'agree' responses, averaged across participants. In each plot are presented the data collected in blocks using random order (in *red*), or ascending order (in *green*), or descending order (in *blue*). Here, ascending refers to blocks for which participants should respond 'agree' at the end of the block (e.g. the *green curves* reads from right to left in the *top panels*, and from left to right in the *bottom panels*)

accepting the color label. Indeed, inflection points were closer to the start of the continuum in ascending blocks (Yellow–Orange: 0.50 vs. 0.54; Green–Blue: 0.48 vs. 0.53). Thus, it appears that with respect to order effects, participants reacted to the negation term included here simply by inverting the polarity of their responses, and still showed an anti-hysteresis effect (Fig 10).

Correlation between positive and negative atomic sentences To shed further light on the similarity between positive and negative atomic sentences, we assessed how behavioral measures such as the location of the inflection point on the color continuum, or the value of the anti-hysteresis effect, were correlated across participants between these two tasks. For the Yellow–Orange continuum, the correlations were significant for both the inflection point ($r = 0.66$, $p < 0.001$) and the anti-hysteresis effect ($r = 0.48$, $p < 0.05$). For the Green–Blue continuum, the correlation for the hysteresis effect was significant ($r = 0.42$, $p < 0.05$), but only for the inflection point it failed to reach significance ($r = 0.24$, $p = 0.25$).

3.2.2 Conjunctive Sentences

Preliminary note Preliminary inspection of individual participants' data revealed that two small sets of observers differed radically from the rest of the cohort, at least

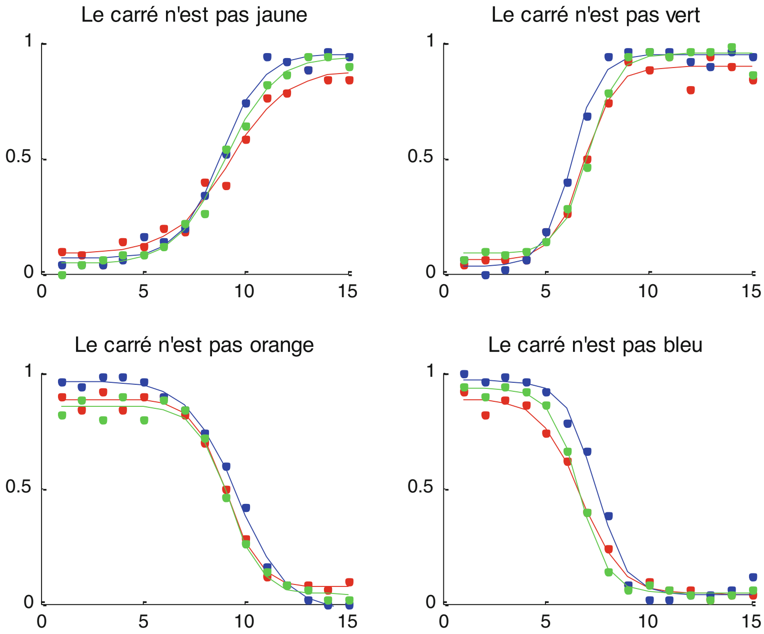

Fig. 9 Raw data for negative atomic sentences x-axis: position on the color continuum, from yellow to orange (*left plots*) or green to blue (right *plots*). y-axis: percentage of 'agree' responses, averaged across subjects. In each plot are presented the data collected in blocks using random order (in *red*), or ascending order (in *green*), or descending order (in *blue*). Here, ascending refers to blocks for which subjects should respond 'disagree', thereby accepting the color label, at the end of the block (e.g. as in the previous figure, the *green curves* reads from right to left in the *top panels*, and from left to right in the *bottom panels*)

during the negative conjunction blocks (e.g. 'The square is yellow and not yellow'). Two observers (S2 and S17) seemed to respond by ignoring the negation part of the sentence (Fig. 11). Three observers adopted a very conservative approach and responded "agree" less than twice on average in these blocks (S4, S14, S25). It is difficult to say that these 3 participants differed qualitatively or merely quantitatively from the rest of the cohort, although in both case it is preferable not to include them with the cohort. Thus 5 participants were excluded from the analyses presented below. We note that these two subgroups have already been reported in previous work involving conjunctive sentences (see Ripley (2011)), who contrasts 'flat' responses and 'slope' responses to the more typical 'hump' responses found in the cohort, see below. Flat responses appear to characterize subjects who systematically reject sentences of the form 'x is P and not P", and slope responses subjects who respond to only one conjunct in such sentences.

The hump effect The first empirical question we asked was whether a conjunction such that "the square is yellow and orange" would be more acceptable than its conjuncts "yellow" and "orange", at the point where the conjuncts cross each other. To investigate this question, we aggregated the atomic data across the three order conditions, then we interpolated the point in the color spectrum at which the conjuncts

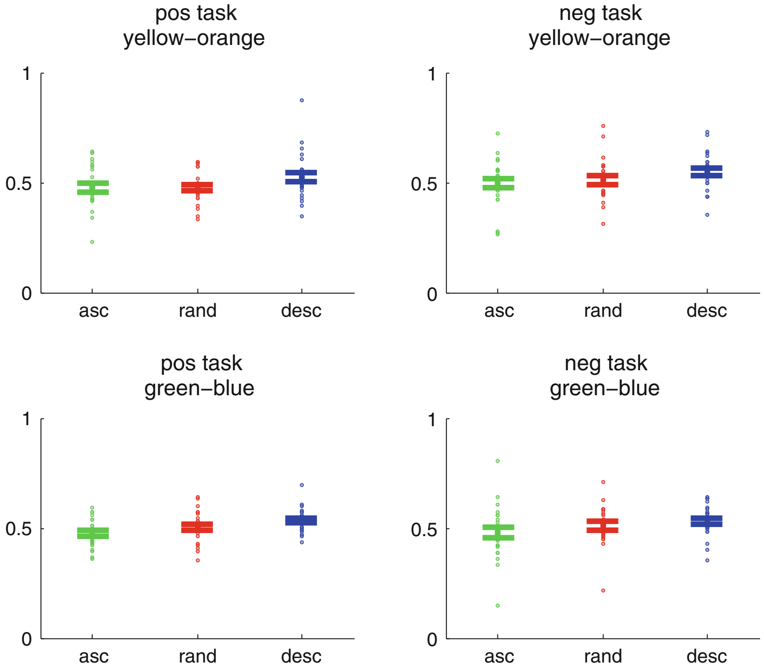

Fig. 10 Position of the inflection points in atomic sentences, for both stimulus sets

are equally acceptable, and calculated the difference between the interpolated degree of acceptation for the conjunction minus the interpolated value for the conjuncts. We call that difference the 'hump effect', meaning that responses for conjunctions make a 'hump' toward the middle of the spectrum (Ripley 2011), and that this hump exceeds the acceptance of both conjuncts. A positive hump effect was first reported in the Alxatib and Pelletier study (2011) in which they note that 32% of subjects who judged true conjunctions of the form '*x* is tall and not tall" also judged false each conjunct separately.

This was done separately, within each color set, for the positive conditions (e.g. comparing "yellow" and "orange" to the average of "yellow and orange" and "orange and yellow") and the negative conditions (e.g. comparing "yellow" and "not yellow" to "yellow and not yellow", but also "orange" and "not orange" to "orange and not orange") (Fig. 14). The two participants that showed a sloping pattern discarded in the previous analysis (S2 and S17) were also discarded in this analysis (although this did not change the results).

T-tests across observers confirmed that the hump effect was significant in most of the conditions we tested. For positive conditions, it was present in both color sets (Yellow–Orange: $t(22) = 2.95$, $p < 0.01$; Green–Blue: $t(22) = 4.43$, $p < 0.001$), and for negative conditions, it was significantly positive for 3 out of 4 conditions ($p < 0.05, t(22) > 2.47$), and not significant only for "yellow and not yellow" ($p = 0.27, t(22) = 1.14$). To compare the amplitude of the hump effect across conditions, we averaged the negative conditions across labels, and then ran an ANOVA

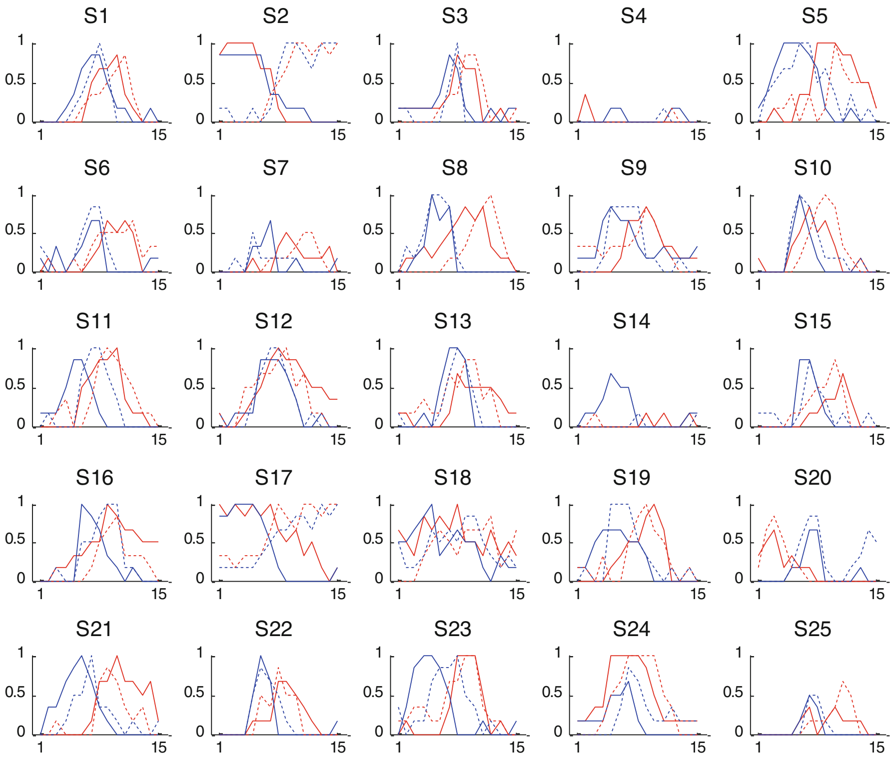

Fig. 11 Individual data for negative conjunctive sentences. x-axis: position on the color continuum, from yellow to orange or from green to blue. y-axis: Proportion of 'agree' responses. Each subplot is an individual observer. Red and blue lines correspond to Yellow–Orange and Green–Blue data, respectively. Full and dashed lines correspond to blocks using different color labels (full red: 'the square is yellow and not yellow', dashed red: 'the square is orange and not orange', full blue: 'the square is green and not green', dashed blue: 'the square is blue and not blue')

across participants with the mode of presentation (positive vs. negative) and the color set (Yellow–Orange vs. Green–Blue) as two within participant factors. This analysis revealed that the amplitude of the hump effect was greater for one stimulus set ($F(1, 22) = 9.08$, $p < 0.01$), with no main effect or interaction from the mode of presentation. On average, the value of the hump effect was 12% for the Yellow–Orange stimulus set and 21% for the Green–Blue stimulus set. Although the amplitude of the hump effect seemed greater for positive conjunctions (20%) than for negative conjunctions (13%), this difference failed to be significant.

Order effects in conjunctive sentences Our second analysis focused on order effects in conjunctive sentences. As the shape of the curve is not sigmoidal, we used the position of the maximum on the response curves (i.e. the position on the x-axis of the maximum, in the curves presented in Figs. 12 and 13), to measure the centre of the "vague region" and assess whether it would be affected by manipulations of presentation order.

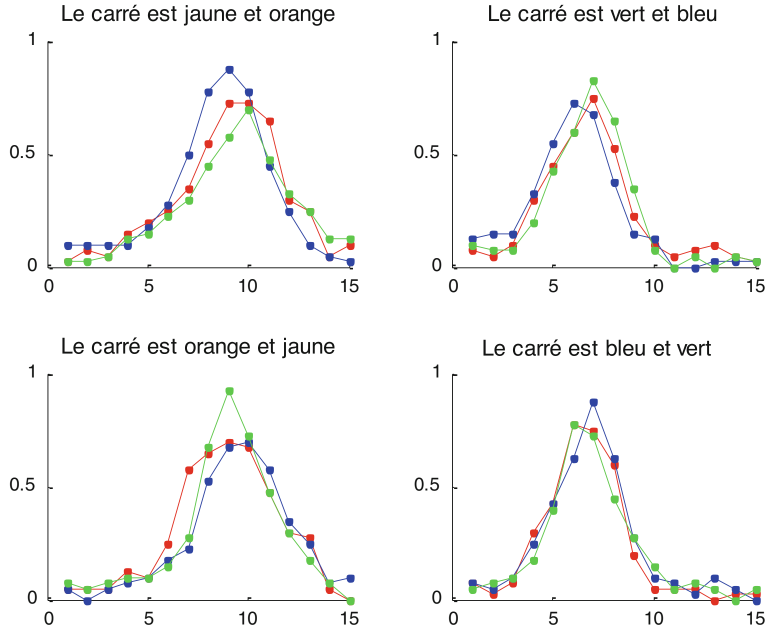

Fig. 12 Raw data for positive conjunctive sentences x-axis: position on the color continuum, from yellow to orange (*left plots*) or green to blue (*right plots*). y-axis: percentage of 'agree' responses, averaged across subjects. In each plot are presented the data collected in blocks using random order (in *red*), or ascending order (in *green*), or descending order (in *blue*). Here, ascending refers to blocks in which the first color label in the sentence should be accepted at the end of the block (i.e. as in previous figures, the *green curves* reads from right to left in the *top panels*, and from left to right in the *bottom panels*), while descending refers to the opposite case

As before, we collapsed across color labels, while holding constant the relation between the presentation order and the labels. We then computed the position of the maximum for each observer and order $\times$ stimulus set conditions, and subjected these values to separate ANOVAs for the positive and the negative conjunctions. These analyses did not reveal any ordering pattern: there were no main effect of order and no interaction between stimulus set and order. Main effects of stimulus set were significant, indicating that the position of the maximum was different between the two color sets, but these effects were not of interest here.

We then considered a second analysis of the same data, by collapsing across labels while holding constant the relation between the presentation order and the physical space, rather than the labels. That is, we make here the assumption that "the square is blue and green" and "the square is green and blue" are equivalent. The position of the maximum was calculated and subjected to separate ANOVAs for the positive and the negative conjunctions. Both ANOVAs indicated significant main effects of stimulus sets. Additionally, for positive conjunctions there was a significant effect of order ($F(2, 38) = 4.65$, $p = 0.015$), indicating that when the presentation order was from blue to green, participants agreed to "the square is blue and green" for shades that were bluer than when the block went from green to blue stimuli.

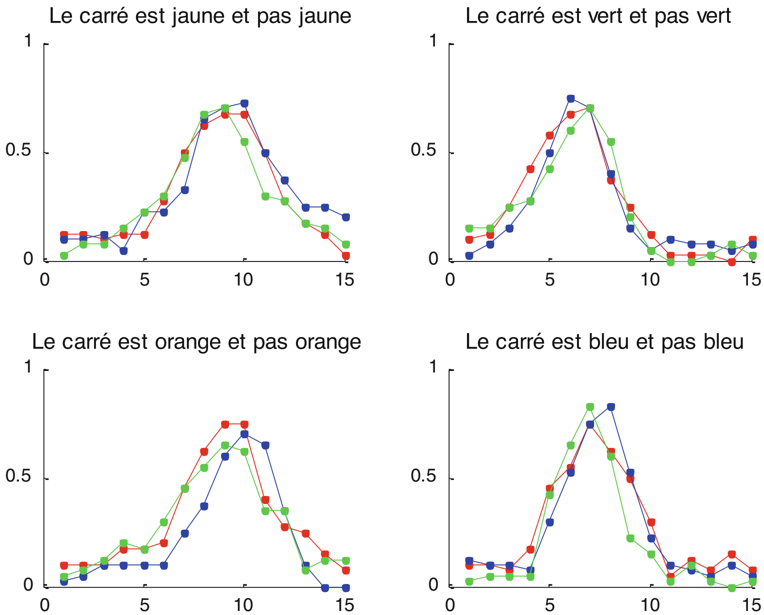

Fig. 13 Raw data for negative conjunctive sentences. Same conventions as in the previous figure

As we did for the atomic sentences, we carried out additional correlation analyses across participants, to assess the similarity between the order effects (measured as the difference between the maxima for one order vs. the other) between positive and negative conjunction blocks. We found a significant correlation for one stimulus set but not for the other one (Yellow–Orange: $r = 0.61$, $p = 0.004$; Green–Blue: *ns*). Finally, we assessed the similarity between conjunction and atomic blocks, with respect to the average order effect across the positive and negative conditions. There, we found significant correlation for both stimulus sets (Yellow–Orange: $r = 0.56$, $p = 0.011$; Green–Blue: $r = 0.53$, $p = 0.015$). These results suggest that although there might be individual variability across observers in their sensitivity to order effects, this variability is consistent across the different modes (e.g. atomic vs. conjunctive) in which order effects have been tested in our experiment.

3.3 Discussion

3.3.1 Negative Hysteresis

The data collected for these blocks constitute a replication of our Experiment 1, with the exception that the response mode changed from selecting one color label out of two ("blue" vs. "green", in Experiment 1) to agreeing vs. disagreeing to a given color label (in Experiment 2). Because of that, the data are indicative that the negative hysteresis found in Experiment 1 does not depend on the comparison between two categories, but that it is equally present for Yes/No judgments relative to a single category.

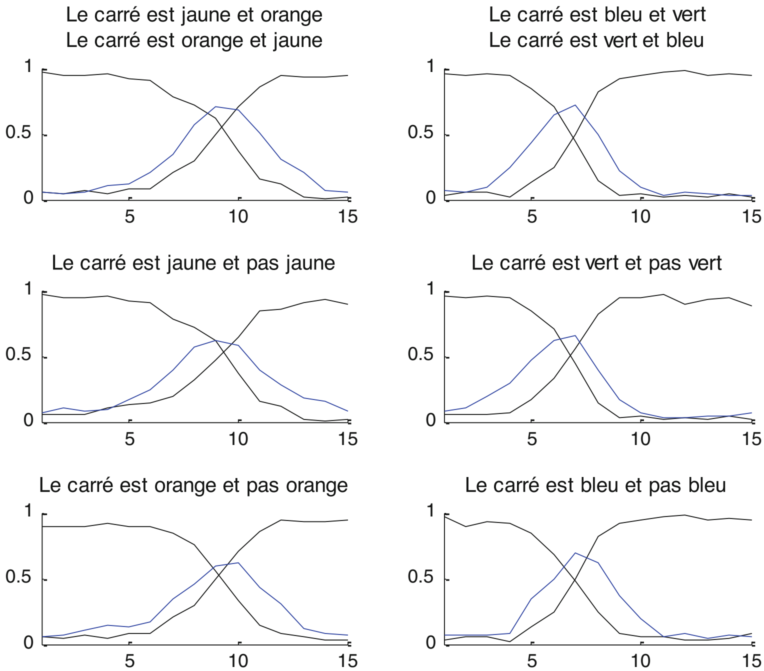

Fig. 14 The hump effect. x-axis: position on the colour continuum, from yellow to orange (*left plots*) or green to blue (*right plots*). y-axis: percentage of 'agree' responses, averaged across subjects. In each plot are presented the data collected in blocks using the conjunction indicated over the plot (in *blue*), or its corresponding conjuncts presented in atomic sentences (in *black*). Positive conditions are presented at the *top*, and negative conditions in the *middle* and *bottom rows*

The explanation proposed for negative hysteresis in the case of Experiment 1 remains applicable in the present case. The reason is that in the strict-tolerant semantics proposed in Cobreros et al. (2012), the negation of a vague predicate too can be asserted in a tolerant and in a strict way. In its strict mode, to say that the square is not blue is to say that the square is clearly not blue. In its tolerant mode, to say that the square is not blue is to say that the square is not clearly blue. The tolerant use of negation remains compatible with the judgment that the square is blue to some extent.[7] As a result, participants who start judging clear blue shades should assent to "the square is blue", again based on principles of accuracy and informativeness. When they reach the first few Blue–Greenish shades, however, they can tolerantly assert "the square

[7] For an atomic sentence relative to a three-valued logic, the negation of a sentence is strictly assertible iff the so-called 'strong' or Gödel negation (mapping False to True and the other values to False) is True; and the negation of a sentence is tolerantly assertible iff the corresponding 'weak' or 'exclusion' negation (mapping True to False and the other values to False) is True; for more complex sentences, however, the negations behave differently: the strict negation is in fact negation under a strong Kleene notion of designated value for assertion, and the tolerant negation negation under an LP notion: both negations are involutive ($\neg\neg A$ and A are equivalent), unlike the Gödel and exclusion negations. We refer the reader to Alxatib and Pelletier (2011) for more details about Gödel and exclusion negation, and to Cobreros et al. (2012) and Cobreros et al. (2014) concerning negation in the strict-tolerant framework.

is not blue", to impart that the square is not of the same blue quality. The opposite holds when participants start from a clear green: this time they should strictly assert that that square is not blue. Once they reach the first noticeably Green–Blueish square, however, it is rational to tolerantly assert "the square is blue".

3.3.2 The Hump Effect

The data involving only atomic sentences in this experiment do not allow us to defeat an epistemicist explanation along the lines suggested for Experiment 1. In principle, participants could choose to use "blue" and "not blue" only for those shades about which they are certain, that is shades standing sufficiently far from some thin underlying boundary. The case of conjunctive sentences allows us to tease the two theories apart, however. The same subjects should indeed choose to disagree with conjunctive sentences such as "the square is blue and green" in a situation in which they are unwilling to assert that the square is blue, and unwilling to assert that it is green. But such agnosticism is not what we find, since participants agree significantly more to "the square is blue and green", and even to "the square is blue and not blue" than they do for either conjuncts in the middle of each color continuum. As pointed out above, this result is exactly consistent with the finding of Alxatib and Pelletier (2011) in the case of static stimuli relative to the vague predicate "tall", and it lends support to the overlap hypothesis we used to account for negative hysteresis in the first place.

3.3.3 Order Effects in Conjunctions

Of the three effects of interest in Experiment 2, order effects in conjunctive sentences are less clear than either the negative hysteresis found in atomic sentences or than the hump effect. We do find a consistent order effect in at least the case of positive conjunctions, but the interpretation of this effect requires some care.

First of all, the predictions of our account regarding order effects in conjunctive sentences are not straightforward. The main question we asked was whether there would be any order effect on conjunctive categories, as opposed to no effect at all. A more refined question is actually whether we could then predict the direction (positive, or negative) or this order effect.

Regarding the first question, our hypothesis was the following: that if borderline cases between green and blue are not a proper region, but only the byproduct of a contextual displacement of a thin boundary relevant for "green" or for "blue", then we should observe no order effect on conjunctions; if, on the other hand, conjunctive categories such as "blue and green", or "blue and not blue" denote a proper region, then we might see order effects in the same way in which we find order effects for the atomic categories "blue", "green" and their negative versions.

Prima facie, the fact that we do observe such order effects may therefore come in support of the overlap view, and go against a purely contextualist view of borderline cases. Upon reflection, however, it is unclear whether our account is really in a position to explain such order effects for conjunctive sentences.

On the account of vague predicates put forward in Cobreros et al. (2012), and endorsed here, "blue and green", like "blue and not blue", can only be consistently interpreted as the conjunction of the tolerant interpretations of each conjuncts (assuming the interpretation to be compositional). The strict interpretation for "blue and green" or "blue and not blue" would indeed be empty, for it would amount to looking for squares that are both clearly blue and clearly green, or clearly blue and clearly not blue, a contradiction. For our conjunctive sentences, therefore, we only have one non-empty interpretation available instead of two, unlike for atomic sentences. As a result, if we found a negative or positive order effect on conjunctive sentences, we would not be able to explain it in the way we explained the order effect found in atomic sentences.

On a different story, "blue and green" might be interpreted (non-compositionally) as denoting a basic category. Suppose that "blue and green" is interpreted as the name of a *sui generis* color (like "turquoise"), then this new category might be susceptible to two interpretations, a strict and a tolerant interpretation, coming with distinct non-empty extensions. In this regard, the fact that we find an order effect in positive conjunctions, but not in negative conjunctions, may be an indication that "blue and green" is more likely to be interpreted as naming a new color than "blue and not blue" is.

Either way, our data regarding order effects in conjunctions are fragile and do not allow us to adjudicate clearly for or against a contextualist view of borderline cases.

4 General Discussion

Our main result in this paper is the negative hysteresis found both in a forced choice task involving two color names and in a YES/NO task involving a single color name. We also observed that this effect does not occur in a task of perceptual matching, and that it is robust to changes of color set. Because of that, we are tempted to conclude that this effect is a proper linguistic effect, pertaining to the semantic representation of color categories in memory. Whether it is specific to the semantic representation of color predicates, or whether it would generalize to arbitrary vague predicates is not something our study can tell for sure, although we conjecture that it should hold of all vague predicates if our proposed explanation based on the strict-tolerant semantic distinction is right.

With regard to the semantic conception of vague predicates, this result is valuable in two respects. Firstly, though predated by Kalmus' earlier finding, this results defeats our initial expectation (also expressed in Kamp 1981; Raffman 1994, 2009; Smith 2008, see fn. 3; Egré 2009) of hysteresis in dynamic sorites. We take it that this expectation is the most common essentially because the regular sorites argument predicts assimilation, rather than contrast, to the category one is coming from.[8] In that sense, the folk expectation of hysteresis may itself be traced to the propensity to reason soritically in the abstract.

Secondly, part of our motivation for testing conjunctive sentences in sorites series was to adjudicate between a contextualist explanation of hysteresis and an overlap view of this phenomenon (see Introduction above). In the face of counterhysteresis, it may seem that both conceptions have to lose ground. Prima facie, the finding of

[8] For a more detailed discussion of contrast versus assimilation in categorization (see Hampton et al. 2005).

an enhanced contrast between connected color categories, like that between a color category and its negation, may appear to run against our proposed view of borderline cases, namely the view on which borderline cases can be characterized in terms of an overlap between distinct categories. Rather, it may be seen as evidence for an underlap view of borderline cases (such as a truth gap view of vague predicates, see again Smith's quote in fn. 3).

The way we view the situation is more complex. What our results indicate is that it would be misguided to identify the overlap view of borderline cases with the prediction of positive hysteresis, and the underlap view with the prediction of an enhanced contrast. What happens, as argued in Egré (2011), is that both for situations of hysteresis and for situations of enhanced contrast, there is a range of cases that subjects equally call "blue" and "not blue", or "blue" and "green", depending on the direction of the transition. In that sense, the finding of an enhanced contrast remains entirely consistent with the characterization of vague predicates in terms of an "overlap". Furthermore, we have independent evidence for the overlap view, based on the fact that subjects agree to sentences of the form "the square is blue and green", or "the square is blue and not blue", in Experiment 2.

Nevertheless, we also have evidence that borderline cases cannot be characterized *exclusively* in terms of an overlap. In the strict-tolerant framework, borderline cases of a vague predicate P are characterized in a dual way, as cases that are tolerantly P and tolerantly not P, and as cases that are neither strictly P nor strictly not P. In Cobreros et al. (2012), we argued that this dual semantic characterization was needed to explain the 'hump' effect for conjunctions first suggested by Alxatib and Pelletier: here we replicated this effect in a dynamic setting. This dual characterization of borderline cases is furthermore consistent with the conceptual space account put forward in Douven et al. (2013), where the borderline region between two color categories can be defined equally as a region of overlap between the extended regions associated to prototypes of two colors, or as the underlap between the restricted regions associated to such prototypes. We see that it is equally relevant to explain the phenomenon of negative hysteresis. Figure 15 gives an abstract representation of the division of the Blue–Green continuum in terms of the strict and tolerant extensions of the predicate "Blue". As illustrated, the strict extension of a vague predicate like "blue" is always a subset of its tolerant extension, and similarly for the strict extension of "not blue" relative to its tolerant extension (we refer to Cobreros et al. 2012 for details). As explained in Sect. 2.3, in an ordered condition from Blue to Green, a switch is predicted at the point where subjects cannot apply their initial predicate strictly anymore.

An additional observation that needs to be made about our protocol is that we did not ask participants to express agreement or disagreement to sentences of the form: "the square is neither blue nor not blue". That option was made available to subjects in Alxatib and Pelletier (2011), who found that subjects assent equally to that description and to the conjunctive description we tested in the present study. In two pilot experiments run prior to this study, we actually probed for positive disjunctive sentences instead, but in a way that was inconclusive. What we did is that we gave our subjects the option to agree or to disagree to sentences of the form "the square is red or not red". Our prediction then was that we might observe a 'hollow' response pattern for such disjunctions, symmetric to the hump response found in conjunctive sentences.

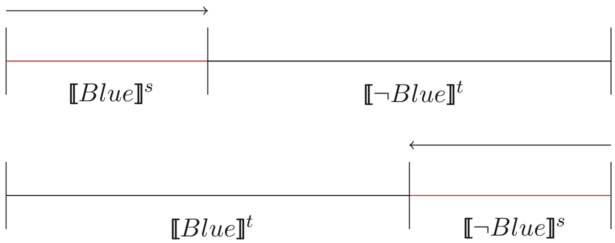

Fig. 15 Negative hysteresis pictured in terms of strict and tolerant extensions. The strict extension of 'blue' and the strict extension of 'not blue' (under the *arrows*, in *red*) underlap, but the tolerant extensions (their complements, in *black*) overlap. *Arrows* indicate the direction of the transition. Switching points occur where strict extensions end. The borderline area between 'blue' and 'not blue' is the overlap (or underlap) region between the switch points

That is, under a strict interpretation of "the square is red or not red", subjects should understand that the square is clearly red or clearly not red, agree to this for clear cases, but disagree in intermediate cases.

In our first pilot study, however, subjects were considerably bothered by such disjunctive sentences and by the fact that they had to respond YES or NO. We conjectured this might be due to the fact that "the square is red or not red" is reinterpreted as an alternative question ("is the square red, or is it not red?"), making such YES/NO answers inappropriate. Because of that, our second pilot study asked subjects to agree or disagree to disjunctions of the form: "the square either is red or is not red", which should not be reanalyzable as alternative questions. Again, however, the data and debriefing indicated that subjects remained very uncomfortable with how to interpret such sentences. Overall, it appears harder or less natural to deny positive disjunctions of the form "*x* is P or not P" for borderline cases than it is to accept such disjunctions under the scope of a negation, as in "*x* is neither P nor not P", or indeed conjunction of negations, depending on which analysis is assumed.

What should be tested, therefore, is whether subjects would assent to 'the square is neither blue nor green' and 'the square is neither blue nor not blue' to the same extent that they assent to 'the square is both blue and green' and 'the square is both blue and not blue" in a dynamic context. Based on Alxatib and Pelletier's findings in a static context, we predict that they should, but the question remains open.

5 Conclusion

We listed three main goals at the outset of this paper. Our first goal was to obtain a precise measure of order effects in the setting of a 'forced march' sorites, and in particular to see whether we would observe a pattern of hysteresis. What we found was indeed an order effect, but the opposite effect, namely an enhanced contrast, in two different tasks, one involving comparison between two color names, the other involving a yes/no categorization relative to one and the same color. As we pointed out, this finding supports the earlier observation by Kalmus (1979) of such an enhanced contrast, but at the same time it undermines a common expectation regarding order effects in sorites series.

Our second goal was to see whether we would find dynamic evidence for an overlap view of vague predicates, rather than evidence for an underlap view, or indeed for a sharp boundary view. We have argued that the effect found is compatible with a dual semantic characterization of borderline cases as cases of overlap and as cases of underlap, in agreement with the account of vague predicates proposed in Cobreros et al. (2012). The hump effect, on the other hand, may be seen as evidence against an epistemicist characterization of borderline cases.

Finally, we sought to clarify the notion of overlap between categories. First, we were interested to compare two notions of overlap, overlap between distinct categories ('blue' vs. 'green'), and overlap between a category and its negation ('blue' vs. 'not blue'). Relative to order effects in atomic sentences, we have not seen a difference between the two notions of overlap in the present setting. Relative to order effects in conjunctive sentences, on the other hand, what we found is some partial evidence for a distinct treatment of 'blue and green' in contrast to 'blue and not blue', but this evidence remains fragile. Secondly, we were interested in the possibility of order effects in conjunctive sentences in order to disentangle the notion of semantic overlap from the notion of contextual shift of boundary between categories (a question initially rooted in the anticipation of positive rather than negative hysteresis). Our evidence does not allow us to draw any firm conclusion here. Importantly, however, we noted that our proposed explanation for order effects in atomic sentences may not directly predict the same order effects in conjunctive sentences, depending on whether such conjunctions are interpreted compositionally or not. Further work is needed to improve our understanding of conjunctions as well as disjunctions of vague predicates more generally.

Acknowledgments This work would not have been possible without institutional support, nor without the help and input of many people. First and foremost, we gratefully acknowledge the Agence Nationale de la Recherche, grant ANR-07-JCJC-0070, program "Cognitive Origins of Vagueness", for funding. We thank Isabelle Brunet, Julien Boyer, Mélanie Brun, Claire Mégnin for their technical assistance, and the LSCP in Paris, where the experiments took place. Warm thanks go to Jean-Michel Hupé, Daniel Pressnitzer, Jean-Luc Schwartz for inspiring discussions and technical advice, as well as to participants in the ANR project MULTISTAP, where different stages of this paper were presented during meetings held in Grenoble, Lyon and Paris. Special thanks go to Diana Raffman for her pioneering work and input on this topic, and for numerous crucial exchanges over the years, as well as to Alan Hájek, Jakub Szymanik, and two anonymous reviewers of the journal for detailed comments and helpful criticisms. We also thank audiences at LSCP, at the Jean-Nicod Institute in Paris, and at ESSLLI 2012 in Opole, where this paper was presented. Further thanks go to various colleagues for related work, conversations and encouragements since 2008, in particular to Pablo Cobreros, Emmanuel Chemla, Igor Douven, James Hampton, Sid Kouider, Robert van Rooij and Jérôme Sackur. We also thank the European Research Council under the European Community's Seventh Framework Program (FP7/2007-2013) and the project 'Borderlineness and Tolerance' (FFI2010-16984), Ministerio de Economía y Competitividad, Government of Spain. Authorship is equally shared between the three authors of this work.

References

Alxatib, S., & Pelletier, F. (2011). The psychology of vagueness: Borderline cases and contradictions. *Mind & Language*, *26*(3), 287–326.

Bonini, N., Osherson, D., Viale, R., & Williamson, T. (1999). On the psychology of vague predicates. *Mind & Language*, *14*(4), 377–393.

Cobreros, P., Egré, P., D., R., & van Rooij, R. (2014). Vagueness, truth and permissive consequence. In D. Achouriotti, H. Galinon, & J. Martinez, (Eds.), *Unifying the philosophy of truth*. Springer (forthcoming).

Cobreros, P., Egré, P., Ripley, D., & van Rooij, R. (2012). Tolerant, classical, strict. *The Journal of Philosophical Logic, 41*(2), 347–385.
Douven, I., Decock, L., Dietz, R., & Egré, P. (2013). Vagueness: A conceptual spaces approach. *The Journal of Philosophical Logic, 42*(1), 137–160.
Egré, P. (2009). Soritical series and Fisher series. In A. Hieke, & H. Leitgeb, (Eds.), *Reduction: Between the mind and the brain* (pp. 91–115). Frankfurt: Ontos Verlag.
Egré, P. (2011). Perceptual ambiguity and the sorites. In R. Nouwen, R. van Rooij, U. Sauerland, & H. Schmitz (Eds.), *Vagueness in communication* (pp. 64–90). Berlin: Springer.
Fine, K. (1975). Vagueness, truth, and logic. *Synthese*, *30*, 265–300.
Hampton, J. A., Estes, Z., & Simmons, C. L. (2005). Comparison and contrast in perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1459.
Hock, H., Bukowski, L., Nichols, D., Huisman, A., & Rivera, M. (2004). Dynamical vs. judgmental comparison: Hysteresis effects in motion perception. *Spatial Vision*, *18*(3), 317–335.
Hock, H., Kelso, J., & Schöner, G. (1993). Bistability and hysteresis in the organization of apparent motion patterns. *Journal of Experimental Psychology*, *19*(1), 63–80.
Horgan, T. (1994). Robust vagueness and the forced-march sorites paradox. *Philosophical Perspectives*, *8*, 159–188.
Hyde, D. (1997). From heaps and gaps to heaps of gluts. *Mind*, *106*(424), 641–660.
Kalmus, H. (1979). Dependence of colour naming and monochromator setting on the direction of preceding changes in wavelength. *British Journal of Physiological Optics*, *32*(2), 1–9.
Kamp, H. (1981). The paradox of the heap. In U. Mönnich (Ed.), *Aspects of philosophical logic* (pp. 225–277). Berlin: Springer.
Kelso, S. (1995). *Dynamic patterns*. Cambridge, MA: MIT Press.
Lindsey, D., Brown, A., & Raffman, D. (2005). Hysteresis effects in a sorites series. Manuscript, cited in D. Raffman 2005.
Raffman, D. (1994). Vagueness without paradox. *The philosophical Review*, *103*(1), 41–74.
Raffman, D. (2005). How to understand contextualism about vagueness. *Analysis*, *65*(3), 244–48.
Raffman, D. (2009). Tolerance and the competent use of vague words. In *Unruly words: A study of vague language*. Chapter 5, Book in Preparation.
Ripley, D. (2011). Contradictions at the border. In R. Nouwen, H.-C. Schmitz, & R. van Rooij, (Eds.), *Vagueness in communication*.
Serchuk, P., Hargreaves, I., & Zach, R. (2011). Vagueness, logic and use: Four experimental studies on vagueness. *Mind & Language*, *26*(5), 540–573.
Smith, N. J. J. (2008). *Vagueness and degrees of truth*. Oxford: Oxford University Press.
Tuller, B., Case, P., Ding, M., & Kelso, J. (1994). The nonlinear dynamics of speech categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(1), 3.
Williamson, T. (1994). *Vagueness*. London: Routledge.