

## Do children know when their room counts as clean?

**Theoretical background** Even if they are harder to acquire than nouns, by the age of 2 y.o., children produce adjectives such as *big*, *little*, *cold*, etc (Blackwell 1998). In order to evaluate whether a particular object, say Dumbo, is big, it is necessary to identify the intended class of comparison, and its relative standard: thus, Dumbo may be small compared to other elephants, but big compared to a mouse. Different studies (Ebeling&Gelman 1998, Smith,Cooney&McCord 1986) highlighted how children are able to rely on contextual factors in order to identify the relevant class of comparison, that can be a normative one (eg. a mitten is big/small relative to the normal size of mittens), a functional one (eg. a dress is big/small for a particular doll) or a perceptual one (e.g. an object is big/small compared to a 2nd object). Moreover, children are able to shift the standard of comparison when required. Some scholars (Kennedy&McNally 2005, Rotstein&Winter 2004;Yoon 1996) recognized the existence of two different kinds of gradable adjectives (GAs, ie. those adjectives that can be “graded”: enter into comparative construction, be modified by degree expressions such as *very*): Relative (or partial) and Absolute (or total) GAs. Rel GAs, eg. *big*, *tall*, *intelligent*, are always evaluated wrt to a standard that is contextually determined; Abs GAs, eg. *full*, *straight*, *clean*, have an intrinsic standard: even if a cloth may be cleaner than another one, what counts as clean is the absence of dirtiness. Kennedy accounted for this fact assuming that GAs project onto ordered scales of degrees; relative GAs activate scales that do not have boundaries; absolute GAs make reference to scales that have an upper and/or lower boundary – and this boundary constitute the intrinsic maximum or minimum standard (max or min std). Thus, *clean* project onto a scale that has an upper end (the total absence of dirtiness – *clean*’s max std), whereas its antonym *dirty* project into the same scale, but with a reverse ordering, and thus it has a lower end that correspond to the min std of dirtiness. Adjectives such as *full/empty* and *open/closed* refer to a scale that is closed on both sides (the max std of *full* is being completely filled; its min std is being completely empty).

**The question** is whether children interpret correctly absolute GAs, that is, whether they know that in order to evaluate whether “This is Adj<sub>Abs</sub>” there is no need to resort to normative, functional or perceptual stds, since the std is intrinsic (thus, a cloth counts as “clean” only if there is no dirt on it, not because, eg, it is cleaner than another cloth close to it).

**Experimental background.** In order to answer this question, Syrett (2007) ran two experiments, whose subjects were 3 y.o., 5 y.o., and adults. She tested 2 relative GAs (*big* and *long*) and 2 absolute GAs (*full*, with a max std, and *spotted*, with a min std). In a Scalar Judgment Task (SJT), subjects were asked to judge if each element in a series of 7 objects that were identical except that they decreased with respect to a relevant dimension (eg. length) had the property denoted by the adjective (they were asked: *Is this Adj?*). In a second experiment, Presupposition Assessment Task (PAT), participants were asked to satisfy the request of a puppet (“Give me the Adj one”), when two objects (sharing the same property) were present. In the case of relative GAs, the request could always be satisfied (eg, subjects were expected to handle the *bigger* object – perceptual std); in the case of absolute GAs, the request was felicitous only if one, and only one, of the objects possessed the property to its intrinsic std. Thus, when asked “give me the *spotted* one”, the request can be satisfied only if there is only one object that is spotted (if both objects have spots on them – even if one has more than the other, the request should be rejected), and when asked “give me the *full* one”, the request can be satisfied only if there is one container that is completely filled (if both containers are filled to some degree, but none is completely full, the request should be rejected). Summing up (and simplifying) the results of both experiments, Syrett found that children had an adult-like behaviour for Rel Gas (even if, in line with previous findings, adults tend to be more categorical) and for *spotted*, the Abs GA with a min std. Unexpected were the findings with *full*: in the SJT, almost all adults judge the 2<sup>nd</sup> element in the series (a container “almost full” of lentils) as “not full”, while 40% of children judge it “full” (and 20% of them continue to assent to “Is this full?” until the 6<sup>th</sup> element, an almost empty container). Also in the PAT, while 88% of adults object to the request “Give me the full one” when there are two not completely full containers, 11 of the 18 children give the puppet the “fuller” container. Syrett noticed that this effect can be partially explained by an influence of the order of presentation, and ran other experiments to further investigate this result, but none of them could offer a convincing explanation of the facts.

**Our experimental study.** We started from Syrett’s unexpected result on *full* and from the hypotheses that could explain it. HYP 1: there is a difference between max std (*full*) and min std (*spotted*) Abs

GAs. HYP 2: there is a difference between Abs GAs that project on scales that have only one end, and those that have two boundaries (recall that *full* projects on a totally closed scale). HYP 3: there is something peculiar on the item chosen for *full* (a container of lentils). HYP 4: the design of the experiment induces a comparative interpretation for *full*. Thus, we compared Abs GAs (1) with a max std (eg. *clean*) vs. a min std (eg. *bent*) – in fact we chose pair of antonyms for both Rel (*big/small*) and Abs (*clean/dirty*) GAs; (2) that project on scales with only one end (eg. *clean*) vs. two ends (eg. *full, open*); (3) for *full/empty* and *open/closed* we used 2 different items with different functions: bottles (that can be “filled”) vs. paint tubes (that can only be “emptied”); and for *open/closed* purses vs. boxes; (4) we designed a new experiment with two sessions. The first part is a Truth-Value-Judgment-Task in which one single object is presented and described by a puppet using an adjective (*This is Adj*); the participant is asked to judge the puppet’s description as “correct”, “incorrect” or “can’t tell/it depends”. The single items had “almost” the property (ie. for Abs GAs they were close to the intrinsic std: an almost full bottle; an almost clean cloth; for Rel GAs they were the 2<sup>nd</sup> item on the series of 7 tested in the SJT – thus the 2<sup>nd</sup> longer rod). The 2<sup>nd</sup> part was a Scalar Judgement Task like Syrett’s one except that we tested 9 scales using a subset of the items of the TVJT.

**Results and discussion.** For a general overview of the results obtained, we report the mean acceptance rate per condition and age in the TVJT in the table below:

TVJT	full (bottle,tube)	empty (bottle,tube)	open (purse, box)	closed (purse, box)	Abs + (eg. clean)	Abs - (eg. dirty)	Rel+ (eg. long)	Rel- (eg. small)
children	88%	44%	100%	50%	13%	87%	75%	61%
adults	93%	15%	100%	46%	4%	93%	56%	65%

We submitted our data to a Person’s Chi-square test (with Yates' continuity correction when required) and Fisher Exact Test by using R. We will focus here only on some relevant findings: (i) considering the type of GAs (Rel. vs. Abs.), we replicated previous findings: both children and adults behave significantly differently in the two conditions, thus distinguishing between these two classes of GAs:  $\chi^2 = 122.14$ ,  $df = 2$ ,  $p < .0001$  for adults;  $\chi^2 = 16.54$ ,  $df = 2$ ,  $p < .0001$  for children; (ii) no difference is instead found comparing positive vs. negative types of adjectives; (iii) interesting results were obtained for the Absolutes projecting on scales closed on both ends. In the first place, differently from what previously found for adults, both children and adults tend to judge “full” something that is only “almost” full, not distinguishing between the type of item used (bottles vs. paint tubes:  $p = .69$  for adults and  $p = .33$  for children, n.s.). However, a marginal difference ( $p = .077$ ) is obtained when comparing children’s acceptance rate of the description “this is full” of the “almost full” bottle in the two tasks: they accept it more in the TVJT than the SJT, as if the presence of a comparison set forced the max std interpretation of *full*, i.e. “completely full”. When the same items are described by the negative adjective “empty”, instead, children behave differently for different items (more “yes” for “almost empty” tubes than bottles,  $p < .001$ ) and differently from adults, that do not differentiate their answers depending on the “container” (adults object to “empty” for both “almost empty” containers,  $p = .47$ , n.s.; children’s acceptance rate differs from adults’ only for paint tubes ( $p < .01$ ), not for bottles ( $p = .47$ , n.s.)). On the contrary, children and adults pattern alike in case of the scale “open/closed”: while they do not differentiate between items (a purse or a box), considering them “open” when “almost totally open”, they behave differently when they are described as “closed”: while they overwhelmingly consider an “almost totally closed” purse as “closed”, they consider an “almost totally closed” box as “non closed” (open),  $p < .0001$ . This difference (between negative and positive poles, or max and min std) is not recorded for scales closed on one end instead. Our findings suggest that: the real difference is between scales closed on one or both ends (*clean* vs. *full*): for the latter, there is also a difference between the direction of ordering (*full* vs. *empty*). This can depend on the semantic representation of totally-closed scales: if both ends constitute the intrinsic std, then an item ought to be judged “full” only if completely full, and “empty” only if completely empty – and there would be no appropriate labels for the middle cases. This suggests that either the semantics for this type of Abs GAs is incorrect, or that it is to be integrated with a different one (with *full* meaning “having a certain amount of fullness”).