



# Convexity and Monotonicity in Language Coordination: Simulating the Emergence of Semantic Universals in Populations of Cognitive Agents

Nina Gierasimczuk<sup>1</sup> · Dariusz Kalociński<sup>2</sup> · Franciszek Rakowski<sup>3</sup> · Jakub Uszyński<sup>4</sup>

Accepted: 12 July 2023 / Published online: 17 August 2023  
© The Author(s) 2023

## Abstract

Natural languages vary in their quantity expressions, but the variation seems to be constrained by general properties, so-called *universals*. Their explanations have been sought among constraints of human cognition, communication, complexity, and pragmatics. In this article, we apply a state-of-the-art language coordination model to the semantic domain of quantities to examine whether two quantity universals—monotonicity and convexity—arise as a result of coordination. Assuming precise number perception by the agents, we evolve communicatively usable quantity terminologies in two separate conditions: a numeric-based condition in which agents communicate about a number of objects and a quotient-based condition in which agents communicate about the proportions. We find out that both universals take off in all conditions but only convexity almost entirely dominates the emergent languages. Additionally, we examine whether the perceptual constraints of the agents can contribute to the further development of universals. We compare the degrees of convexity

---

✉ Nina Gierasimczuk  
nigi@dtu.dk

Dariusz Kalociński  
dariusz.kalocinski@gmail.com

Franciszek Rakowski  
francis.rakowski@gmail.com

Jakub Uszyński  
uszynskijakub@gmail.com

- <sup>1</sup> Department of Applied Mathematics and Computer Science, Technical University of Denmark, Richard Petersens Plads, 2800 Kgs. Lyngby, Denmark
- <sup>2</sup> Institute of Computer Science, Polish Academy of Sciences, ul. Jana Kazimierza 5, 01-248 Warsaw, Poland
- <sup>3</sup> Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw, ul. Tyniecka 15/17, 02-630 Warsaw, Poland
- <sup>4</sup> National Information Processing Institute, Al. Niepodległości 188 B, 00-608 Warsaw, Poland

and monotonicity of languages evolving in populations of agents with precise and approximate number sense. The results suggest that approximate number sense significantly reinforces monotonicity and leads to further enhancement of convexity. Last but not least, we show that the properties of the evolved quantifiers match certain invariance properties from generalized quantifier theory.

**Keywords** Quantity terminologies · Quantifiers · Language evolution · Approximate number sense · Agent-based modelling · Generalized quantifier theory · Semantic universals · Monotonicity · Convexity

## 1 Introduction

Natural languages include a variety of quantity terms, among them numerals (e.g., *one*, *two*, *three*) and quantifiers (e.g., *at least two*, *a few* or *half of*). Such words are important in communication—every language includes, at minimum, expressions for denoting initial natural numbers (Gordon, 2004; Pica et al., 2004). Typically, numerical or, more generally, quantitative terminologies are well-developed, which is documented by the considerable range of such systems across cultures (Menninger, 1969; Bach et al., 1995; Matthewson, 2008; Keenan & Paperno, 2012).

Despite cross-linguistic differences, the issue of so-called *linguistic universals*, properties shared by (almost) all natural languages, is a recurring topic. Such properties have been studied in phonology (Hyman, 2008), syntax (Greenberg, 1963; Chomsky, 1965; Croft, 1990), and semantics (Lindsey & Brown, 2009; Jäger, 2010). In this paper we focus on convexity and monotonicity—two out of many candidates for quantifier universals studied in applied mathematical logic (Barwise & Cooper, 1981; Peters & Westerståhl, 2006). Convexity and monotonicity can be easily explained in the case of colors and scalar terms, but their general principles remain the same across domains. A colour term, such as *red* or *blue*, is associated with a region of cognitive *colour space*. Convexity of a region means, figuratively, that one can enclose it tightly with a rubber band, and that enclosure will contain no gaps. A gradable adjective, such as *tall* or *cold*, is associated with a region of the linear scale of degrees (of *tallness* or *coldness*), similar to an interval. Such a region is monotone if its membership extends to any greater (or lesser) degree, as it is conveniently expressed by the following: if *X* is tall (cold), then anything taller (colder) than *X* is also tall (cold). The convexity universal with regard to colour terms states that the meanings of *naturally occurring* simple colour terms are convex (Jäger, 2010). Similarly, the monotonicity universal with regard to gradable adjectives states that meanings of *naturally occurring* simple gradable adjectives are monotone (Carcassi, 2020). The notions of convexity and monotonicity of quantity terms, and the corresponding universals, are translations of the above ideas to more abstract spaces of numerosities.

In the first language evolution model of quantifiers (Pauw & Hilferty, 2012), meanings were convex by default, which excluded the possibility of explaining the emergence of convexity. We allow that universals are not directly hardwired in cognition. Where do they come from? One of the possible answers is that their source lies in general constraints of human cognition: linguistic constructions that are ill-adapted to

such constraints tend to be eliminated in favor of the more fitting ones (Christiansen & Chater, 2016). It has been shown, for example, that quantifier meanings satisfying certain universals, including monotonicity, are simpler in terms of (approximate) Kolmogorov complexity (van de Pol et al., 2019) and learnability (Steinert-Threlkeld & Szymanik, 2019). The same acquisition mechanism favours monotone quantifiers in artificial iterated learning which consists of repeated production-acquisition events, arranged in a chain (Carcassi et al., 2019). Quantifiers resulting from such experiments are sometimes unlike those we encounter in natural languages. This might be, in part, due to the fact that constrained cognition is not the only pressure to which languages are sensitive. This drawback has been somewhat mitigated in another study where, using a different language evolution model than (Carcassi et al., 2019), quantifiers evolving under the competing pressures for simplicity and informativeness become more natural (Steinert-Threlkeld, 2019).

Existing explanations of quantity universals do not take into account horizontal (synchronic) dynamics of languages or, more specifically, language coordination, i.e., the mechanisms that work at the level of interaction between users, and that drive the adjustment of users' word-meaning mappings. The closest to these dynamics is the iterated learning model (Carcassi et al., 2019) but again—this model reflects only vertical (diachronic) dynamics where a previously learned language is passed on to the next generation. In contrast, in this paper we use one of the state-of-the-art solutions to the language coordination problem, originally applied to colour terms (Steels & Belpaeme, 2005). The language coordination model provides a hypothetical mechanistic explanation of the transition from the stage in which a modelled population does not have a language for communicating about a certain domain, to the stage in which it has one. Our main modification of the model consists in using number (instead of colour) stimuli, and therefore redefining the perception of agents.

The first question we ask is the following:

Do semantic universals, such as convexity and monotonicity, emerge as a result of language coordination? (Question 1)

Apart from being detached from horizontal dynamics, existing explanations of universals with regard to quantity terms are implicitly based on the assumption that the perception of quantities by cognitive agents is precise. This assumption is sound in small contexts involving 1–5 objects. Typically, however, such a restriction is unrealistic. In the context of larger quantities, humans use a separate magnitude estimation system—approximate number sense (ANS)—and produce an instant perception of quantity at the cost of accuracy, with an error proportional to the intensity (cardinality) of the perceived input (Dehaene, 1997). This ability is actually more fundamental than precise counting, which develops only later in childhood (Cantlon et al., 2006), including more sophisticated numerical concepts (Feigenson et al., 2004).

The second question we examine in this paper is the following:

Do the constraints of ANS on the level of individuals support the emergence of semantic universals, such as convexity and monotonicity? (Question 2)

To address the above questions we employ a methodology of computer simulations. Models based on similar principles have been used extensively in language evolution studies (see, e.g., Steels, 2012).

Our experiment concerns a population of agents in which a certain interaction protocol is repeated many times. Each step of the simulation can be described as a combination of two levels: a single- and a multi-agent one.

On the single-agent level, each agent perceives quantitative stimuli according to an activation pattern (coded by what we call a *reactive unit*). Here the experimental conditions vary—we equip our agents with either ANS-based or exact reactive units. The exact reactive unit allows precise recognition of the numerosity given in the stimulus. For ANS-based reactive units a larger estimation error can occur more likely as the stimulus size increases. The application of quantity-based reactive units is what sets our model apart from Steels and Belpaeme (2005). The rest of the model remains the same. The agent groups individual reactive units into categories which are then used to discriminate between stimuli in a so-called *discrimination game* (see Sect. 3.1 for details). Depending on the result of discrimination, the agent updates its repertoire of categories accordingly in order to enhance her chances of successful discrimination in the future. The structure of this game, and of reactive units (designed to reflect the constraints of ANS), is important for Question 2.

On the multi-agent level, the agents interact in randomly chosen pairs, according to the rules of the so-called *guessing game*. In each such interaction one of the agents is assigned the role of the *speaker* and the other one is the *hearer*. They are jointly shown two stimuli, of which one is selected to be the *topic*.<sup>1</sup> The speaker perceives both stimuli and, by employing the discrimination game, finds a category that distinguishes the topic from the other stimulus. At this point, the speaker reaches out to her individual (weighted) associative map between categories and terms/words (we often refer to this mapping as her language or terminology), finds a term that best corresponds to that category in her language and utters that word. The hearer then looks up the word uttered by the speaker and the category that has the strongest binding to that word in his language. He makes a guess by pointing to the stimulus with the highest response in that category. The coordination game is successful if the hearer points to the topic. In this case the language associations responsible for that guess are reinforced, others are decreased. In the case the game fails, the associations responsible for the guess are decreased. This game implements horizontal, i.e. between-agent, dynamics, and therefore is especially important for Question 1.

The paper is structured as follows. The perception model and the language evolution model are described in Sects. 2 and 3, respectively. In Sect. 4 we specify our notions of *meaning*, and the two semantic universals of *convexity* and *monotonicity*. We also relate these concepts to generalized quantifier theory. Section 5 describes our experiments aimed at comparing the degrees of convexity and monotonicity of terminologies resulting from the cognitively-plausible ANS-based model with the respective degrees obtained from a cognitively-neutral, ANS-independent model. Section 6 concludes the paper.

---

<sup>1</sup> Perhaps this is another slight difference from the model by Steels and Belpaeme (2005) where agents were presented four stimuli.

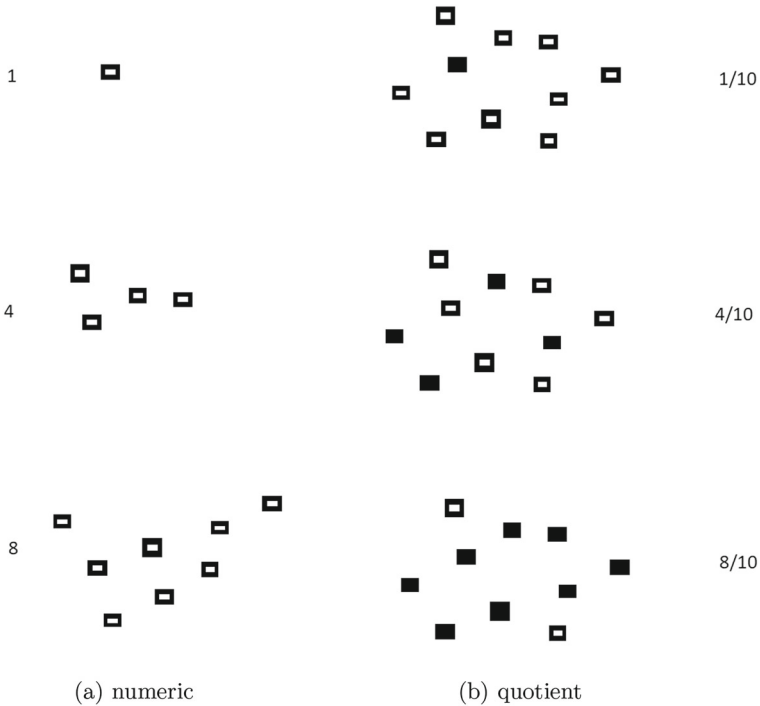


Fig. 1 Types of stimuli (a) numeric and (b) quotient

## 2 Perception Module

An agent can receive two types of number stimuli: numeric or quotient.

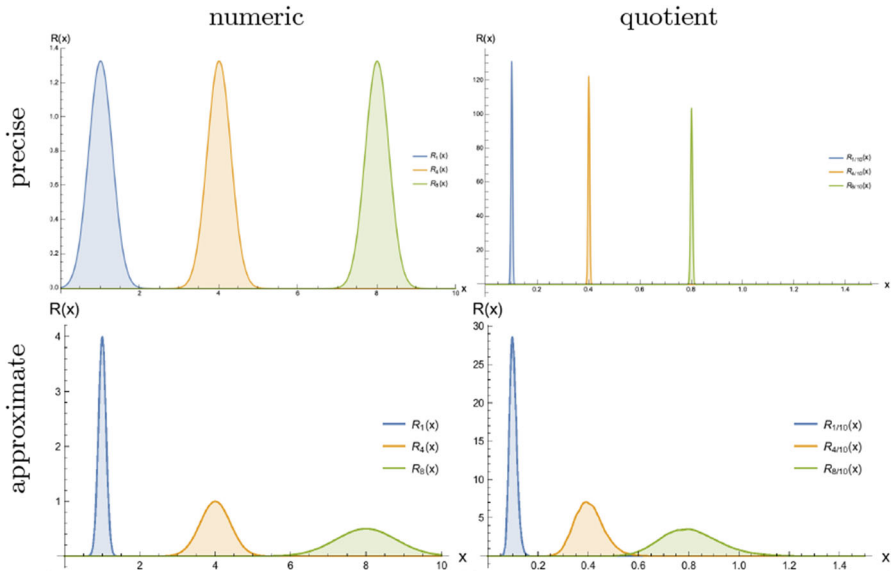
A *numeric stimulus* is a positive integer. It can be interpreted as the number of objects presented to an agent (Fig. 1a).

A perceived stimulus can be mistaken for a different but close-enough quantity. We capture this using a vague representation formalism of *reactive units* (defined by Steels and Belpaeme (2005) for the special case of colour space). Here, a reactive unit is a bell-shaped real function centered on a certain quantity and fading with the distance from it. The reactive unit for the numeric stimulus  $n$  has the following form:

$$R_n \sim \mathcal{N}(n, \sigma^2), \tag{1}$$

where  $\mathcal{N}(n, \sigma^2)$  is the normal distribution with mean  $n$  and some standard deviation  $\sigma$ .

We distinguish two modes of perception: precise and approximate. Precise number sense allows for accurate discrimination between any two distinct stimuli. We capture it by taking a sufficiently small constant standard deviation ( $\sigma = 1/3$ ) across all stimuli so that, for each  $n$ , almost all of  $R_n$ 's mass is wrapped around  $n$  (top of Fig. 2). The second mode, approximate number sense (ANS), incorporates scalar variability,



**Fig. 2** Reactive units of numeric-based stimuli 1, 4, 8 (left) and quotient-based stimuli 1/10, 4/10, 8/10 (right) according to precise ( $\sigma = 0.3$ , top row) and approximate ( $\sigma = n \cdot 0.1$ , bottom row) number sense

or Weber’s law, which makes vagueness grow linearly with the magnitude of the stimulus (Fechner, 1966):

$$R_n \sim \mathcal{N}(n, (n \cdot 0.1)^2). \tag{2}$$

The above model of ANS is inspired by signal detection theory (Green & Swets, 1966) and has received both experimental and theoretical support (Pica et al., 2004; Halberda & Feigenson, 2008; Cheyette & Piantadosi, 2020). Agents using precise number sense respond to stimuli as if they were capable of precise counting regardless of the overall size or intensity of the stimulus.

*Quotient stimulus* is a (reduced) fraction  $\frac{n}{k}$ . It can be interpreted as the ratio of  $n$  objects of a given kind among all  $k$  objects presented to an agent (Fig. 1b). Quotient stimuli lack systematic cognitive modelling literature. For example, the question whether ANS is recruited in human ratio processing seems to be open (O’Grady et al., 2016; O’Grady & Xu, 2020). Here, we decided to follow a hypothesis formulated in the most recent study (O’Grady & Xu, 2020) which suggests that ratio processing utilizes number processing by combining representations of the numerator and the denominator into the corresponding ratio distribution. Given reactive units  $R_n, R_k$ , representing numerical stimuli  $n, k$  respectively, the reactive unit for the quotient stimulus  $\frac{n}{k}$  is given by the following ratio distribution:

$$R_{n/k} \sim R_n/R_k. \tag{3}$$

Two properties of this model are worth mentioning. First is that, assuming the approximate mode of perception, we clearly see an analogue of scalar variability, or Weber's law (Fig. 2, bottom-right). This is good because similar effects have been reported in experiments with human subjects (O'Grady & Xu, 2020).

The second property of the above model is that for any two equal ratios (e.g.  $9/23 = 27/69$ ) the corresponding ratio distributions are identical ( $R_{9/23} = R_{27/69}$ ). This property can be easily derived from known analytical representations of the ratio distribution  $X/Y$  of two independent normal random variables  $X, Y$  (see, e.g., Equation 1 in Díaz-Francés and Rubio, 2013). There, two additional requirements are imposed on  $X, Y$ —positive means and coefficients of variation less than one. These requirements are satisfied by ratio distributions considered in this paper: each  $R_n \sim \mathcal{N}(n, n \cdot 0.1)$  has positive mean and the coefficient of variation is  $\delta = (n \cdot 0.1)/n < 1$ . This property justifies our restriction to use only reduced fractions in our simulations.

The above model of ratio perception implies that the perception of proportions like  $10/10, 7/7$  etc. is still characterized by a substantial variance. It seems that in such cases the non-proportional elementary quantifier *all* is empirically used (which seems to be consistent with the results of empirical results in Szymanik and Zajenkowski, 2010). Therefore, we restrict the set of stimuli to reduced fractions from the open interval  $(0, 1)$ .

In the following, if the exact form of a reactive unit is irrelevant, we denote it by  $z$ , possibly with subscripts. When we write  $z(x)$ , we mean  $p_z(x)$ , where  $p_z$  is the probability density function associated with  $z$ .

## 2.1 Categories

While a reactive unit describes the cognitive response to a single stimulus, a category may include several such units, collectively representing a set of stimuli. Formally, a category is defined as a linear combination of  $k$  ( $k \in \mathbb{N}$ ) reactive units  $z_1, z_2, \dots, z_k$ , and corresponding non-negative weights  $w_1, w_2, \dots, w_k$  (describing the significance of each unit within the category):

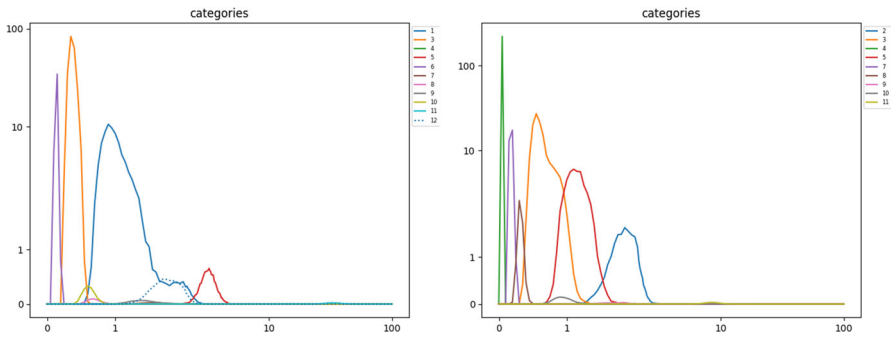
$$c(x) := \sum_{i=1}^k w_i z_i(x). \quad (4)$$

Suppose that an agent equipped with a category  $c$  observes a stimulus  $s$ .  $c$  generates a response reflecting how well  $s$  fits into the category  $c$ . Formally, the response is given by the scalar product of the category  $c$  and the reactive unit  $R_s$ :

$$\langle c | R_s \rangle := \int c(x) p_{R_s}(x) dx, \quad (5)$$

where  $p_{R_s}$  is the probability density function associated with  $R_s$ .<sup>2</sup>  $\langle c | R_s \rangle$  may be regarded as the measure of similarity between  $c$  and  $R_s$ . The more overlap between  $c$

<sup>2</sup> We note that the integrability of  $c \cdot R_s$  follows from the integrability of reactive units and basic facts from measure theory (Halmos, 1976).



**Fig. 3** Examples of emerged categorizations for two different agents at the same step of a test trial. The agent on the right currently has nine categories. The category marked in orange is likely based on at least two reactive units, while the one marked in grey has a small weight and will most likely fade away with time

and  $R_S$ , the greater the value of  $\langle c|R_S \rangle$  and, clearly,  $\langle c|R_S \rangle$  is close to 0 if the overlap is negligible. Also,  $\langle c|R_S \rangle$  is modulated by the weights of the reactive units of  $c$ . These properties provide some theoretical support for the choice of scalar product as a measure of how well a stimulus fits into a category.<sup>3</sup>

To illustrate the above concept of categories, we run a test trial involving 10 agents and plotted categories of two different agents after 99 steps (see Fig. 3).

### 3 Language Coordination Model

The coordination model consists of agents, each having her own *language*. A language is a weighted associative map between words (which happen to be some randomly generated strings) and categories (see Sect. 2.1). Agents engage in two kinds of games: a *discrimination game* (Steels, 1997) and a *guessing game*. In a discrimination game, played individually, an agent observes a random pair of stimuli, called *context*, of which one is the *topic*. The agent uses her categories to distinguish the topic from the remaining stimulus. This is done by checking whether the category with the strongest response to the topic is different from the analogical category for the remaining stimulus. Such a category is called the winning category. If no winning category is found, the agent updates her repertoire of categories, either by adding a new category centered on the topic or extending one of the existing categories.

In a guessing game, two randomly picked agents (designated as *speaker* and *hearer*) meet. They observe a random context but the topic is known only to the speaker. The speaker plays the discrimination game. Upon success, she finds a word with the strongest association with the winning category in her language, and utters that word. The hearer then looks up the uttered word in his language, finds the category that has the strongest binding with that word and points to the stimulus with the highest response in that category. If the hearer correctly points to the topic, the game is successful,

<sup>3</sup> In the model by Steels and Belpaeme (2005), a response of  $c$  to a colour stimulus  $\bar{x}$  is  $c(\bar{x})$ . This is justified because  $\bar{x}$  stands for a complex activation pattern reduced to a single point in the CIELab space.



and the strength of the corresponding bindings within their languages is increased, and other bindings are decreased. If the game fails, the strength of the corresponding bindings are decreased.

In a single run of the model, agents engage in thousands of such guessing games, building up their repertoires of categories and individual languages.

### 3.1 Discrimination Game

An agent is presented with a context  $(q, q')$ , with one of the stimuli,  $q_T \in \{q, q'\}$ , being the topic.

Let  $C$  be the set of categories of the agent.

1. If  $C = \emptyset$ , i.e., if the agent (as of yet) has no categories to use, the game fails.
2. If  $\langle c|R_q \rangle = 0$  or  $\langle c|R_{q'} \rangle = 0$  for all  $c \in C$ , i.e., if all categories the agent is equipped with give a null response to perceptions of at least one stimulus, the game fails.
3. Let  $c = \underset{c \in C}{\operatorname{argmax}} \langle c|R_q \rangle$  and  $c' = \underset{c \in C}{\operatorname{argmax}} \langle c|R_{q'} \rangle$ , i.e.  $c$  and  $c'$  are the categories which for the perception of the stimuli  $q$  and  $q'$ , respectively, generate the strongest response.

(a) If  $c \neq c'$ , the game is successful and the winning category ( $c$  if  $q = q_T$ ,  $c'$  otherwise) is returned.

(b) If  $c = c'$ , the discrimination game fails.

Discriminative success (DS) describes how well an agent performs in a series of such games. Let  $ds_j^a = 1$  if the  $j$ -th discrimination game of an agent  $a$  is successful. Otherwise, let  $ds_j^a = 0$ . The cumulative discriminative success of an agent  $a$  at game  $j$  for the last  $n$  games (we use  $n = 50$ ) is:

$$DS(n)_j^a := \frac{1}{n} \sum_{i=\max(0, j-n)}^j ds_i^a. \quad (6)$$

In other words, interpreting discriminative games outcomes as observations, DS might be understood as an accuracy metric restricted to the last 50 samples, where “true value” is a discriminative success. If  $a, j$  are known/irrelevant, we write  $DS$  to mean  $DS(50)_j^a$ . In other words, it is an accuracy metric (see James et al., 2013), Section 2.2.3) restricted to the last 50 samples.

#### Adjustment in the Case of Success

When a discrimination game is successful, the output category  $c$  is reinforced to enhance its chances of winning similar games in the future. Let  $w_i, z_i, i = 1, 2, \dots, k$ , be the weights and reactive units of  $c$ , respectively. Reinforcement is computed in the following way:

$$w_i := w_i + \beta \cdot \langle z_i|R_{q_T} \rangle, \quad (7)$$

i.e., the weights of  $c$  are increased proportionally to the corresponding reactive units' responses to the topic. In our experiment we take  $\beta$ , the learning rate, to be 0.2.

### **Adjustment in the Case of Failure**

The set of categories is adjusted to avoid a similar failure in the future.

1. Suppose  $C = \emptyset$ . We set  $C = \{c\}$ , where  $c$  is a new category with one reactive unit  $R_{q_T}$  and weight 0.5.
2. Suppose  $\langle c | R_q \rangle = 0$  or  $\langle c | R_{q'} \rangle = 0$  for all  $c \in C$ . A new category is added in the same way as in the first point above (for  $q$ , if  $q$  does not yield any positive response, otherwise for  $q'$ ).
3. Suppose  $c = c'$ . If  $DS < \delta$ , we set  $C := C \cup \{c\}$ , where  $c$  is a new category with one reactive unit  $R_{q_T}$  and weight 0.5. If  $DS \geq \delta$ , the category  $c$  (and thus  $c'$ ) is adapted by appending  $R_{q_T}$  with initial weight 0.5.  $\delta$  is called the discriminative threshold and it is set to 95%.

After the completion of the game and the associated adjustments, the weights of all reactive units of all categories of the agent are decreased by a small factor  $0 < \alpha < 1$  ( $\alpha = 0.2$ ),  $w_i = \alpha w_i$ , which results in slow forgetting of categories. If all weights of a category are smaller than 0.01, the category is deleted.

## **3.2 Language**

In our model the *language* provides a bridge between words from an agent's lexicon and the external stimuli, via categories (and as a consequence also the reactive units). An important feature of words is that they can be directly transferred between agents, while categories and reactive units cannot. An agent's lexicon is a dynamic object—initially it is empty but it can grow or shrink with time. Technically speaking, words in the lexicon are short strings (we use the package `gibberish` to generate nonsensical pronounceable random strings, see <https://pypi.org/project/gibberish/>).

Let  $F$  be a lexicon and  $C$  a set of categories. Language is a function  $L : F \times C \rightarrow [0, \infty)$  which assigns a non-negative weight  $L(f, c)$  to each pair  $(f, c) \in F \times C$ .  $L(f, c)$  indicates the strength of the connection between the word  $f$  and the category  $c$ . If  $L$  is a language, we denote its lexicon by  $F(L)$  and its set of categories by  $C(L)$ . The changes in the language are caused both by additions and deletions performed on the lexicon and categories, and by changes of the strengths of connections between words and categories in the language.

To illustrate the above concept of language, we plotted two representations of individual languages (see Figs. 4, 5) based on the test trial used to produce Fig. 3.

## **3.3 Guessing Game**

Language is used and adjusted in a guessing game between two agents, the speaker and the hearer. They are given a common context consisting of two stimuli; as before, one of them is distinguished as the topic. Crucially, only the speaker is aware which stimulus constitutes the topic. The speaker's goal is to utter a word so that the hearer will be able to guess the topic correctly.

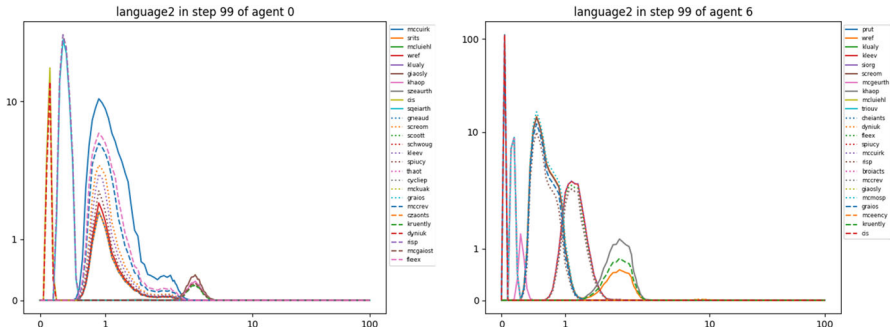


Fig. 4 Illustration of languages of two different agents at the same step of a simulation. Within each language, the competition between different words associated with the same categories is clearly visible

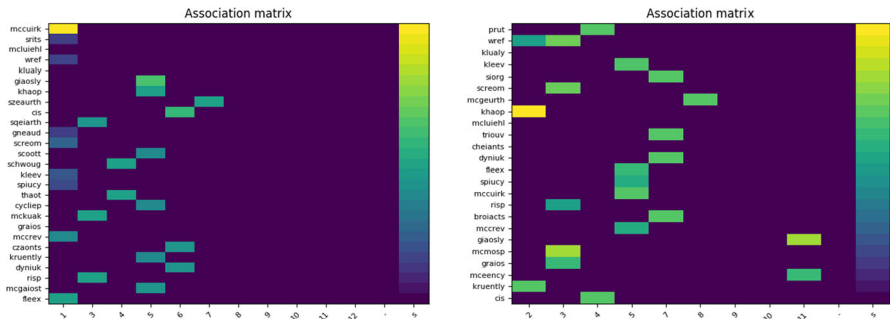


Fig. 5 The matrices show the strength of connections between words and categories in the languages from Fig. 4

Let  $D_S, C_S, L_S$  be the lexicon, the categories and the language of the speaker. Similarly, we have  $D_H, C_H, L_H$  for the hearer.

1. The speaker and the hearer are presented with a common context consisting of two stimuli:  $q_1, q_2$ . One of the stimuli, denote it by  $q_S$ , is the topic. The information whether  $q_S = q_1$  or  $q_S = q_2$  is available only to the speaker.
2. The speaker tries to discriminate the topic  $q_S$  from the rest of the context by playing the discrimination game (Sect. 3.1). If a winning category is found (denoted by  $c_S$ ), the game continues, otherwise the game fails.
3. The speaker searches for words  $f$  in  $D_S$  which are associated with  $c_S$ , i.e., such that  $L_S(f, c_S) > 0$ .

If no associated words are found (i.e.,  $L_S(f, c_S) = 0$ , for all  $f \in D_S$ ) or  $L_S$  is empty, the speaker creates a new word  $f$  (i.e.,  $D_S := D_S \cup \{f\}$ ), sets  $L_S(f, c_S) = 0.5$  and utters  $f$ .

Now suppose that some associated words are found. Let  $f_1, \dots, f_n$  be all words associated with  $c_S$ . The speaker chooses  $f$  from  $f_1, \dots, f_n$  such that  $L_S(f, c_S) \geq L_S(f_i, c_S)$ , for  $i = 1, 2, \dots, n$ , and conveys  $f$  to the hearer. The choice of such a word is consistent with the concept of *pragmatic meaning*, defined later in this section.

4. The hearer looks up  $f$  in her lexicon  $D_H$ .  
 If  $f \notin D_H$ , the game fails and the topic is revealed to the hearer. The repair mechanism is as follows. First, the hearer adds  $f$  to her lexicon (i.e.,  $D_H := D_H \cup \{f\}$ ). Next, the hearer plays the discrimination game to see whether or not she has a category capable of discriminating the topic. If one is found, say  $c$ , the hearer creates an association between  $f$  and  $c$  with the initial strength of 0.5 (i.e.,  $L_H(f, c) = 0.5$ ).  
 Suppose the hearer finds  $f$  in  $D_H$ . Let  $c_1, c_2, \dots, c_k$  be the list of all categories associated with  $f$  in  $L_H$  (i.e.,  $L_H(f, c_i) > 0$  for  $i = 1, 2, \dots, k$ ). The hearer chooses  $c_H$  from  $c_1, \dots, c_k$  such that  $L(f, c_H) \geq L(f, c_i)$  for  $i = 1, 2, \dots, k$ . The hearer points to the stimulus, denoted by  $q_H$ , that generates the highest response for  $c_H$  (i.e.,  $q_H = \operatorname{argmax}_{q \in \{q_1, q_2\}} \langle c_H | R_q \rangle$ ).
5. The topic is revealed to the hearer. If  $q_S = q_H$ , i.e., the topic is the same as the guess of the hearer, the game is successful. Otherwise, the game fails.

#### **Adjustment in the Case of Success**

The speaker increases the strength between  $c_S$  and  $f$  by  $\delta_{inc}$  and decreases associations between  $c_S$  and other word forms by  $\delta_{inh}$ . The hearer increases the strength between  $c_H$  and  $f$  by  $\delta_{inc}$  and decreases the strength of competing words with the same category by  $\delta_{inh}$ .

#### **Adjustment in the Case of Failure**

The speaker decreases the strength between  $c_S$  and  $f$  by  $\delta_{dec}$ . The hearer decreases the strength between  $c_H$  and  $f$  by  $\delta_{dec}$ .

In the experiment we take  $\delta_{inc} = \delta_{inh} = \delta_{dec} = 0.2$ .

## **4 Convexity and Monotonicity**

### **4.1 Meaning**

Let us first define the notion of *meaning* in our model. Intuitively, the meaning of a word is the set of stimuli associated with it in the language. In this general notion of meaning we do not take into account contextual factors and other competing words whose meaning might occupy overlapping areas of the perceptual space. However, we can also think of a more restrictive concept of *pragmatic meaning* of a word, in which contextual factors and interactions with other words play a role. Pragmatic meaning of a word is then a subset of the meaning of that word—the set of stimuli which have the *strongest binding* with that word in the language.

Let  $Q$  be the set of stimuli. The *meaning* of a word  $f$  in a language  $L$  is defined by:

$$[f]^L := \{q \in Q : \sum_{c \in C(L)} L(f, c) \langle c | R_q \rangle > 0\}. \quad (8)$$

In words, a stimulus  $q$  contributes to (is an element of) the meaning of  $f$  if some category  $c$  associated with  $f$  gives a positive response to  $q$ .

A justification for (8) comes from the guessing game. Let  $q, q'$  be a context with the topic  $q$ . The speaker utters  $f$ . The hearer searches for the category with the highest association to  $f$  and finds  $c$ . To choose  $q$  and win the game, it suffices that  $\langle c | R_q \rangle > \langle c | R_{q'} \rangle$ . In principle, any stimulus  $q$  generating a positive response to  $c$  can outperform (in terms of the response magnitude) some other stimulus  $q'$ . Hence, the interpretative process on the part of the hearer suggests that  $q$  indeed contributes to the meaning of  $f$ .

Now we turn to the notion of *pragmatic meaning* which depends on the context and interactions with other words:

$$[f]_p^L := \{q \in Q : \exists c [f \in \underset{L(f',c)>0}{\operatorname{argmax}}_{f' \in F(L)} L(f', c), c \in \underset{(c|R_q)>0}{\operatorname{argmax}} \langle c | R_q \rangle]\}. \quad (9)$$

In words,  $q$  is an element of the pragmatic meaning of  $f$  if  $f$  is maximally (and positively) associated with  $c$  that gives a maximal (and positive) response to  $q$ . The former condition corresponds to between-word interaction while the latter one corresponds to context-dependence.<sup>4</sup>

## 4.2 Convexity

The notion of *convexity* has paved its way into cognitive research thanks to the theory of conceptual spaces (Gärdenfors, 2000, 2014). According to this theory, meanings can always be represented geometrically, and natural categories must be convex regions in the associated conceptual space. There exists evidence supporting this theory for logical words like quantifiers (see, e.g., Barwise & Cooper, 2010; Keenan & Stavi, 2019).

Convexity of a set in a conceptual space is usually defined using an in-betweenness relation. In our case, the strictly ordered set is the set of stimuli with the *less than* relation,  $(Q, <)$ . The set  $S \subseteq Q$  is *convex* if for any  $a, b, c \in S$ : if  $a < b < c$ , then  $b \in S$ . Hence, the (pragmatic) meaning of  $f$  in  $L$  is convex if it is a convex set in the space of stimuli  $Q$ . The above definition of convexity of meaning is sufficient for our purposes and can be derived from a more subtle analysis of convexity for quantifiers presented by Chemla et al. (2019).

## 4.3 Monotonicity

The notion of *monotonicity* plays a prominent role in the traditional model-theoretic semantics (Peters & Westerståhl, 2006). It is often listed among quantifier universals; it is believed to hold for the denotations of all naturally occurring simple determiners

<sup>4</sup> It is worth mentioning that our notion of pragmatic meaning is closely related to the concept of exhaustification (for a recent overview consult Trinh, 2019). Moreover, explicating meaning in terms of strength of connections is closely related to the discussion of prototype-based semantics of quantifiers, (see Tiel et al., 2021).

(see, e.g., Barwise & Cooper, 1981; Keenan & Stavi, 1986). Consider the following sentences:

Many good chess players know advanced tactics. (10)

Many good chess players know tactics. (11)

(11) follows from (10). The only difference in the above sentences is that *advanced tactics* is replaced by the more inclusive expression *tactics*. The inference would remain valid if we replaced any expression *A* with a more general expression *B*. This shows that the validity of the above inference is entirely based on the meaning of the quantifier *many*. The property of meaning that guarantees that is called (upward) monotonicity. In other words, a quantifier is upward monotone if whenever it can be truthfully applied to a set of objects, it can be truthfully applied to its supersets.

Another variant of monotonicity is called downward monotonicity. Consider the following sentences:

Few beginning chess players know tactics. (12)

Few beginning chess players know advanced tactics. (13)

Sentence (13) follows from (12) and the only difference is that *tactics* is substituted by a more specific term *advanced tactics*. The inference would be valid if we replaced any expression *B* with a more specific expression *A*. We call a quantifier monotone if it is either upward or downward monotone.

We attribute monotonicity to meanings of words in our model. The justification of this is that a pragmatic meaning represents only the fragment of the overall meaning—the one that is most contextually and linguistically salient. For example, even though *most* can be truthfully used in situations where all objects possesses a given property, it then might make more sense to use *all* instead.

To pinpoint the notion of monotonicity in our model, let  $f$  be a word in language  $L$ . We say that  $[f]^L$  is *monotone* if it is upward closed with respect to  $\leq$  (i.e., if  $q \in [f]^L$  and  $q \leq q'$  then  $q' \in [f]^L$ ) or downward closed with respect to  $\leq$  (if  $q' \in [f]^L$  and  $q \leq q'$  then  $q \in [f]^L$ ).

In mathematical logic, quantity terms are studied in the field of generalized quantifier theory (GQT). The (universal) properties of quantifiers, such as monotonicity, are central to that study. It is then natural to ask: what is the place of our notions of monotonicity and convexity within GQT? We answer this question below. The reader who would rather skip the theoretical analysis and move on to the experimental part is invited to jump directly to Sect. 5.

#### 4.4 Monotonicity and Convexity in GQT

Generalized quantifier theory (GQT, Lindström 1966, Mostowski 1957, see also Peter & Westerståhl, 2006 for an overview) is dedicated to a mathematical study of quantifiers, with a special focus on their semantics. The purpose of this section is to clarify what properties of generalized quantifiers are in the scope of this paper, bringing our

work closer to the classical analysis of quantity expressions offered by mathematical logic.

In GQT, the meaning of a quantifier expression, such as ‘Some’, ‘All’, ‘At least five’, or ‘An even number of’, is often taken to be a class of (finite) models which satisfy that quantifier.<sup>5</sup> Below we will restrict our attention to finite models of type  $\langle 1 \rangle$ , i.e., finite models of the form  $M = (U, A)$ , where  $U$  is the universe (e.g., the domain of dots), and  $A \subseteq U$  corresponds to the scope (e.g., the predicate ‘black’).<sup>6</sup> A quantifier can then be viewed as a characteristic function over the class of finite models—each finite model either belongs to the quantifier or it does not.

**Definition 1** A quantifier  $Q$  of type  $\langle 1 \rangle$  is a class of models of type  $\langle 1 \rangle$ . If  $M = (U, A) \in Q$ , we will also write  $Q_M(A)$ .

Note that our present study (and also GQT in general) is about an isolated class of expressions that concern *only* quantities. They disregard possible non-quantitative features, like the spatial arrangement of objects, or their proper names. In other words, quantifier expressions are *topic-neutral*. The logical value of a quantifier expression in a finite model depends then solely on the *sizes* of the relevant subsets. This condition is mathematically expressed as closure under isomorphism, ISOM.

**Definition 2** [ISOM] A quantifier  $Q$  of type  $\langle 1 \rangle$  satisfies ISOM iff for any two models  $M_1$  and  $M_2$  with their respective universes  $U_1$  and  $U_2$  and any  $A_1 \subseteq U_1$  and  $A_2 \subseteq U_2$ , we have that if  $\text{card}(A_1) = \text{card}(A_2)$  and  $\text{card}(U_1 - A_1) = \text{card}(U_2 - A_2)$ , then  $Q_{M_1}(A_1) \Leftrightarrow Q_{M_2}(A_2)$ .

Given ISOM, any model of type  $\langle 1 \rangle$  relevant for the interpretation of a quantifier expression can be reduced to a pair of numbers  $(k, n)$ , such that  $\text{card}(U - A) = k$  and  $\text{card}(A) = n$ .

**Definition 3** Let  $Q$  be an ISOM type  $\langle 1 \rangle$  quantifier. The *quanti-relation*  $Q \subseteq N \times N$  corresponding to the quantifier  $Q$  is defined in the following way: for any  $k, m \in N$ ,  $Q(k, m)$  if and only if there is a model  $M = (U, A)$ , such that  $\text{card}(U - A) = k$ ,  $\text{card}(A) = m$ , and  $Q_M(A)$ .<sup>7</sup>

This allows representing the space of relevant finite models as the so-called *number triangle* (van Benthem, 1986), where each model  $M = (U, A)$ , up to isomorphism, is identified with a pair of numbers  $(k, n)$ , such that  $\text{card}(U) = k + n$  and  $\text{card}(A) = n$ , see the left-hand side of Fig. 6. In the triangle the  $n$ -th row enumerates all possible finite models of size  $n$ , again, up to isomorphism. A quantifier (via its quanti-relation) can then be represented geometrically as a *shape* in this space—the right-hand side of

<sup>5</sup> Note the similarity to our empirical notion of meaning, defined to be a set of stimuli that have positive association with the expression.

<sup>6</sup> In the applications of GQT to natural language the prevalent approach is to interpret quantifier expressions as quantifiers of type  $\langle 1, 1 \rangle$ , requiring models to be of the form  $M = (U, A, B)$ , where  $A$  is the set of dots, and  $B$  is the set of, e.g., black objects. Our observations can easily be extended to that domain, by restricting the domain of  $\langle 1, 1 \rangle$  quantifiers to the special subclass: the so-called CE quantifiers, which lends itself easily to a similar analysis (Peters & Westerståhl, 2006).

<sup>7</sup> Note that for clarity we use sans-serif font to denote the quanti-relation  $Q$  corresponding to the quantifier  $Q$ .

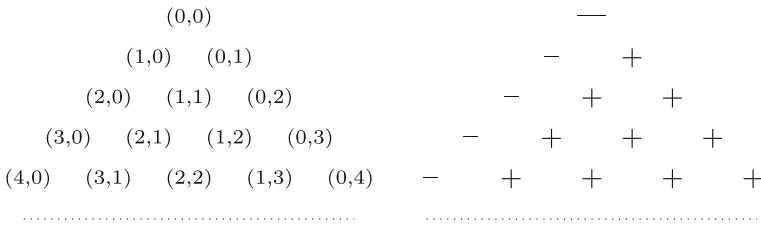


Fig. 6 The number triangle and the representation of ‘At least 1’

Fig. 7 A fragment of the number triangle showing the relevant directions of monotonicity

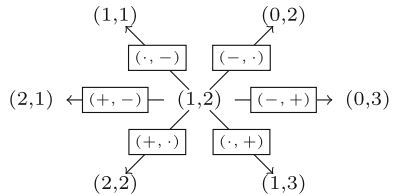


Fig. 6 shows the quantifier ‘At least 1’, with the models satisfying the quantifier marked with a ‘+’. ‘At least 1’ includes all the finite models, where the second parameter is equal to, or larger than 1, i.e.,  $A \neq \emptyset$ .

The number triangle can be seen as a simple precursor for a *conceptual space for numerosities*. Its ever growing base accounts for the growing variety of the possible distributions of elements over the parts of the model, as the overall number of elements increases. The immediate neighbours of a finite model are those obtained by increasing or decreasing the parameters by one.

The six directions specified by arrows in the Fig. 7 can be seen as six kinds of monotonicity. In classical GQT we distinguish two kinds of monotonicity in the right argument (the horizontal direction in Fig. 7), and four kinds of monotonicity in the left argument (the diagonals in Fig. 7).

The most familiar kind of monotonicity is monotonicity in the right argument, i.e., closure under horizontal steps.

- ( $\rightarrow$ ) if  $Q(k, n)$  and  $k \neq 0$ , then  $Q(k - 1, n + 1)$
- ( $\leftarrow$ ) if  $Q(k, n)$  and  $n \neq 0$ , then  $Q(k + 1, n - 1)$

Note that these are not the ‘easiest’ steps to take—they require a simultaneous expansion of one of the constituents and reduction of the other one. They do, however, have a natural interpretation: if a model is fixed (and so is its size), this kind of monotonicity tells us that the truth of a quantifier expression is preserved under taking subsets or supersets. If the sentence ‘At least one car is speeding’ is true in the model, then the sentence ‘At least one car is moving’ is also true in that model, as the set of moving objects is a superset of the set of speeding objects. The properties ( $\rightarrow$ ) and ( $\leftarrow$ ) correspond to upward right-monotonicity ( $\text{MON}\uparrow$ ) and downward right-monotonicity ( $\text{MON}\downarrow$ ), respectively.

The other kinds of monotonicity, coded by the diagonal transitions in the number triangle, are obtained by altering one of the parameters while leaving the other one unchanged.



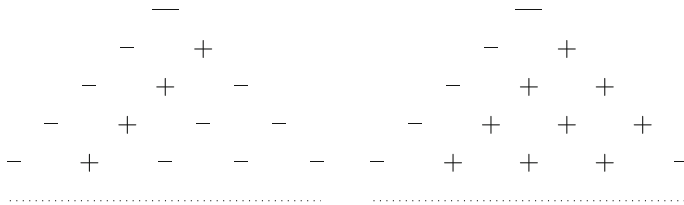


Fig. 8 ‘Exactly 1’ and ‘Between 1 and 3’

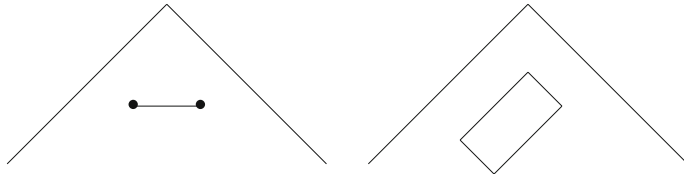


Fig. 9 R- CONT and CONT

- (↘) if  $Q(k, n)$  then  $Q(k, n + 1)$
- (↙) if  $Q(k, n)$  then  $Q(k + 1, n)$
- (↗) if  $Q(k, n)$  and  $k \neq 0$ , then  $Q(k - 1, n)$
- (↖) if  $Q(k, n)$  and  $n \neq 0$ , then  $Q(k, n - 1)$

Some combinations of these properties have natural interpretations. For instance, the property of upward left-monotonicity,  $\uparrow$ MON (so-called persistence), combines two south-pointing arrows (↙) and (↘). Any quantifier of the form ‘At least  $n$ ’ (see Fig. 7) is persistent—once the threshold of  $n$  elements is met, adding an element to any part of the model will not make the quantifier false. Another important property of quantifiers, symmetry, is a combination of (↙) and (↗); as a special case of symmetric quantifiers we want to distinguish the expression ‘Exactly  $n$ ’, which will be our candidate for exact counting term, see Fig. 8, left-hand side.

Finally, smoothness is the combination of (↗) and (↘). Note that smoothness is very ‘contagious’: it spreads vertically and horizontally, entailing the property ( $\rightarrow$ ). The quantifier ‘At least half’ is smooth, while the quantifier ‘At most half’ is co-smooth, i.e., it is its reverse—the combination of (↖) and (↙).

Let us try to approach the concept of convexity in a similar way. A good starting point is *continuity*, which, as with monotonicity, can be interpreted horizontally (relative to a given model) or vertically (with respect to a growing universe).

**Definition 4** [R- CONT] A type  $\langle 1 \rangle$  quantifier  $Q$  is *right-continuous* if and only if  $Q_M(A_1)$ ,  $Q_M(A_3)$ , and  $A_1 \subseteq A_2 \subseteq A_3$ , implies  $Q_M(A_2)$ .

In words, right-continuity means that if two models of the same cardinality satisfy a quantifier, then the models that are located between the two in the number triangle also satisfy the quantifier. In terms of quanti-relations: let  $Q(n_1, k_1)$ ,  $Q(n_2, k_2)$ ,  $n_1 + k_1 = n_2 + k_2 = \ell$ , then  $Q(n, k)$  for  $n + k = \ell$  such that  $\min(k_1, k_2) \leq k \leq \max(k_1, k_2)$ .

A more general notion of continuity accounts for both arguments.

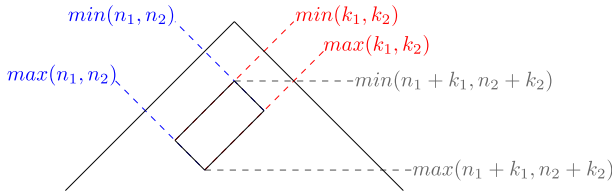


Fig. 10 Parameters in Definition 5 (of CONT)

**Definition 5** [CONT] Let  $Q$  be an ISOM quantifier of type  $\langle 1 \rangle$ , and let  $Q$  be its quanti-relation.  $Q$  is *continuous* if  $Q(n_1, k_1)$  and  $Q(n_2, k_2)$  implies that  $Q(n, k)$  for  $n, k$  such that:

1.  $\min(n_1 + k_1, n_2 + k_2) \leq n + k \leq \max(n_1 + k_1, n_2 + k_2)$ ;
2.  $\min(n_1, n_2) \leq n \leq \max(n_1, n_2)$ ;
3.  $\min(k_1, k_2) \leq k \leq \max(k_1, k_2)$ .<sup>8</sup>

This condition says that the quantifier is true in all models belonging to the quadrangle spanning between any two models making the quantifier true, see Fig. 10. Intuitively, CONT could be a good candidate for an adequate notion of convexity, which assumes the distances and in-betweenness of this conceptual space.

With these notions in mind, let us now consider the expressions and properties that we study in our experiments in the two different conditions, the numeric and the quotient stimuli.

### Numeric Stimuli

Since numeric stimuli concern only the number of one kind of elements, the corresponding quantifiers only depend on the set  $A$ , and not on the set  $U - A$ . Quantifiers of this kind satisfy the property called *extension*, EXT, sometimes also referred to as *domain-independence*.

**Definition 6** [EXT] A quantifier  $Q$  of type  $\langle 1 \rangle$  satisfies EXT iff for any two models  $M_1$  and  $M_2$  with their respective universes  $U_1$  and  $U_2$ , if  $A \subseteq U_1$  and  $A \subseteq U_2$ , then  $Q_{M_1}(A) \Leftrightarrow Q_{M_2}(A)$ .

This means that, assuming ISOM, the numeric expressions will correspond to quanti-relations satisfying  $(\sphericalangle)$  and  $(\sphericalcap)$ , so-called symmetric quantifiers. The exact counting is then expressed by ‘Exactly  $n$ ’, see left-hand side of Fig. 8.

A numeric expression will be called monotonically increasing if its quanti-relation satisfies  $(\rightarrow)$ , and so the quantifier is  $\text{MON} \uparrow$ . The conditions of symmetry and  $\text{MON} \uparrow$  together give us (a subclass of) persistent quantifiers. An example of such a quantifier is ‘At least 1’, see Fig. 6. Symmetric numerical quantifiers that are also  $\text{MON} \downarrow$  (their quanti-relation satisfying  $(\leftarrow)$ ), such as ‘At most 1’, coincide with the so-called co-persistent quantifiers. To sum up, the quantifiers underlying the semantics of numeric expressions in our experiments are EXT and SYMM.

In line with our definitions in Sects. 4.2 and 4.3, the meaning of a numerical expression is monotone, if it is  $\text{MON} \uparrow$  or  $\text{MON} \downarrow$ . We will call it *convex* if it is R-CONT.

<sup>8</sup> For an illustration of the meaning of these parameters see Fig. 10.

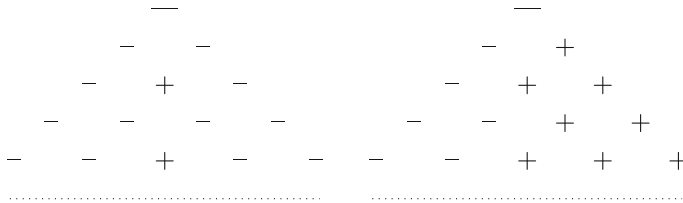


Fig. 11 ‘Exactly half’ and ‘At least half’

It requires little effort to realize that SYMM and R-CONT together imply CONT. Note that right-monotonicity together with SYMM implies CONT, but there are CONT quantifiers that are not right-monotone, for instance *Between 1 and 3*, see Fig. 8 on the right. Interestingly, they can be recovered as intersections (conjunctions) of a monotonically increasing and a monotonically decreasing numerical quantifier.

**Quotient Stimuli**

The quantifiers adequately describing quotient stimuli are still ISOM, but are not EXT, since their truth value depends on both  $card(A)$  and  $card(U - A)$ . First, let us propose a definition of an exact quotient quantifier, with an example, ‘Exactly half’, depicted in Fig. 11 (left).

**Definition 7** A quantifier  $Q$  is *quotient-exact* if there are two non-negative integers  $p$  and  $q \neq 0$ , such that for any model  $M = (U, A)$ ,  $Q_M(A)$  iff  $\frac{card(A)}{card(U)} = \frac{p}{q}$ .

The exact recognition of proportions gives a very fragmented picture. Such quantifiers are only true in some models at levels with universes divisible by the denominator of a given reduced fraction. Unlike the exact numerical quantifiers, the quotient-exact ones are not SYMM and they are not CONT. The question is then how to formally treat their monotonicity. In our experiment, quotient expressions are deemed monotonically increasing if their truth-value extends to larger proportions, like in the case of the Rescher quantifier ‘At least half’. This corresponds to the condition of smoothness: ( $\nearrow$ ) and ( $\searrow$ ) together,<sup>9</sup> For the monotonically decreasing quotient-sensitive quantifiers like ‘At most half’, we would require co-smoothness.

There is an alternative, perhaps more natural, way to approach monotonicity of proportional quantifiers. Let us start by introducing the concept of *quotient-sensitivity*, which requires that the quantifier has the same truth-value in models that enjoy the same proportion of elements.

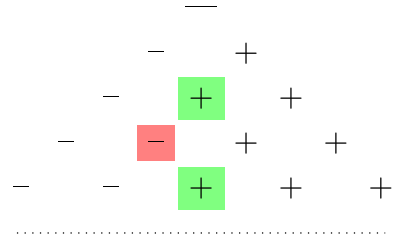
**Definition 8** A quantifier  $Q$  is *quotient-sensitive* if and only if for any two models  $M_1 = (U_1, A_1)$  and  $M_2 = (U_2, A_2)$ , if  $\frac{card(A_1)}{card(U_1)} = \frac{card(A_2)}{card(U_2)}$  then  $Q_{M_1}(A_1)$  iff  $Q_{M_2}(A_2)$ .

The proofs of the following two propositions can be found in the Appendix.

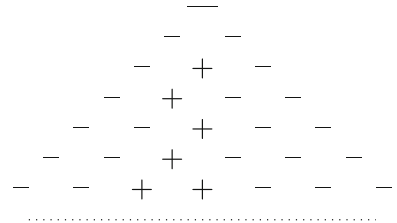
**Proposition 1** Let  $Q$  be a quantifier of type  $\langle 1 \rangle$ . If  $Q$  is quotient-sensitive, then  $Q$  is ISOM.

<sup>9</sup> Note that unlike in the case of numerical quantifiers, extending the quotient-exact to the right, ( $\rightarrow$ ), will not give us quantifiers like *At least half* since it will miss the levels of the number triangle where the total number of elements is odd.

**Fig. 12** ‘At least half’ is not CONT



**Fig. 13** ‘Between one third and a half’: an intersection of the smooth quantifier ‘At least one third’ and the co-smooth quantifier ‘At most half’



**Proposition 2** *Let  $Q$  be a quotient-sensitive quantifier of type  $\langle 1 \rangle$ .  $Q$  is  $\text{MON}\uparrow$  iff  $Q$  is smooth.*

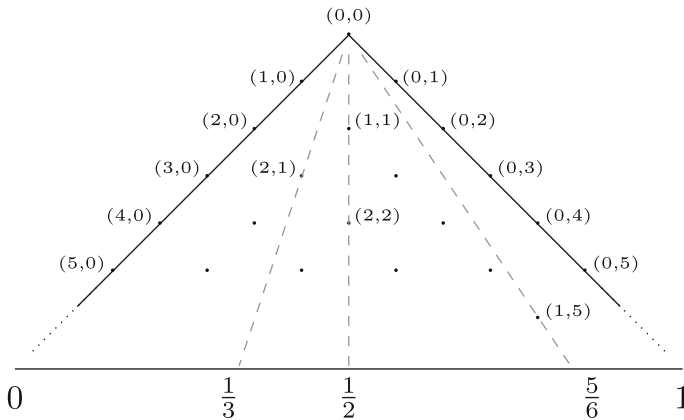
For the quotient-based stimuli, we restrict to a subclass of ISOM, which we call quotient-sensitive quantifiers. All quotient-exact quantifiers are obviously quotient-sensitive. They correspond to expressions like ‘Exactly half’. To talk about proportional monotone quantifiers, we simply impose the condition  $\text{MON}\uparrow$ , which the quotient-sensitive class gives us the desired semantics of expressions like ‘At least half’. ‘At most half’ and can be obtained by requiring  $\text{MON}\downarrow$ .

Finally, let us focus on convexity for quantifier expressions. First note that quotient-sensitive monotone quantifiers are R-CONT, but do not satisfy the more general condition of CONT—a counterexample is given in Fig. 12.

In our experiments for quotient stimuli, we take convexity to mean that if an expression is true in two models with different proportions of elements, then it is also true in all the models with proportions in-between those two. This corresponds to expressions of the kind ‘Between one-third and a half’ (see Fig. 13). They coincide with intersections (conjunctions) of a  $\text{MON}\uparrow$  and  $\text{MON}\downarrow$  quotient-sensitive quantifiers. The result is R-CONT, but not CONT.

The notion of R-CONT concerns only the right argument of the quanti-relations. The question now is if it can adequately describe convexity of expressions whose truth-value clearly depends also on the other argument. Note that our restriction to quotient-sensitive expressions ‘flattens’ the number triangle space into a linearly ordered set of rational numbers (see Fig. 14). Under this interpretation the one-dimensional R-CONT seems justified.

To conclude, the expressions emerging in numerical stimuli experiments are those with the semantics corresponding to EXT and SYMM type  $\langle 1 \rangle$  quantifiers. Among those: monotone are  $\text{MON}\uparrow$  or  $\text{MON}\downarrow$  (they coincide with persistent and anti-persistent, respectively), and the convex ones are R-CONT (such also happen to be CONT). The expressions emerging in quotient stimuli experiments coincide with quotient-sensitive quantifiers. Among them the monotone ones satisfy  $\text{MON}\uparrow$  or  $\text{MON}\downarrow$  (which we prove



**Fig. 14** Projection of the number triangle pairs on the number line. Each line that starts from  $(0, 0)$  and goes through a pair  $(k, n)$  has the property that it intersects every other pair  $(k', n')$  that yields the same proportion (i.e.,  $\frac{n}{k+n} = \frac{n'}{k'+n'}$ ), and determines the point on the number line that can be labelled with the reduced fraction  $\frac{n}{k+n}$

to be equivalent to smoothness and co-smoothness), and the convex ones are R-CONT, but not CONT. This defines the scope of our simulation experiments in the context of generalized quantifier theory.

### 5 Experiments

In our experiments, a numeric-based stimulus is sampled from the uniform discrete probability distribution  $\mathcal{U}\{1, 20\}$ ; a quotient-based stimulus is obtained by first sampling a denominator  $k \sim \mathcal{U}\{1, 20\}$  and then a non-greater numerator  $n \sim \mathcal{U}\{1, k\}$ .

Since  $R_n, R_k$  are Gaussian distributions, we could, in principle, compute the probability density function of  $R_n/k$  based on known analytical expressions (see, e.g., Hinkley, 1969; Pham-Gia et al., 2006). In practice, we resort to an empirical approximation of  $R_n/k$ .<sup>10</sup>

The code for experiments and data visualization is written in Python and Mathematica. Experiments are based on parameters mentioned earlier in the text. A simulation consists of 30 trials. Within a trial, 10 agents evolve languages across 3000 steps. At each step, agents are paired randomly to play a guessing game. There are separate simulations for numeric-based and quotient-based stimuli, as well as for the condition where agents are equipped with the ANS and one where they are not. This gives us a total of 4 experimental conditions. First, we show that the model defined in Sect. 3 is valid, i.e., that it provides a solution to the language coordination problem in the domain of quantities. Next, we check the extent to which convexity and monotonicity are represented in languages emerging across the different conditions.

<sup>10</sup> We first draw two equally-sized samples  $(n_i)_{i=1}^s$  and  $(k_i)_{i=1}^s$  from  $R_n$  and  $R_k$ , respectively. Then we derive a density histogram based on ratios  $(\frac{n_i}{k_i})_{i=1}^s$ . Finally, we obtain a desired approximation by interpolating the resulting histogram.

## 5.1 Model Validity

We use three types of metrics to evaluate validity: discriminative success, communicative success, and number of active terms in the lexicon. Similar measures have been used in previous studies (Steels & Belpaeme, 2005; Pauw & Hilferty, 2012). Discriminative and communicative success are usual accuracy metrics (James et al., 2013).

The cumulative discriminative success of a population  $A$  for the last  $n$  games is defined as:

$$CDS(n)_j := \frac{1}{|A|} \sum_{a \in A} DS(n)_j^a, \quad (14)$$

where  $DS(n)_j^a$  is defined in (6), as before we use  $n = 50$ .

The measure of communicative success,  $CS$ , reflects the percentage of successful linguistic interactions in a population. The requirement is that at some point  $CS$  will get to 50% and stay above that threshold later on.<sup>11</sup> Formally, let  $cs_j^a = 1$  if the  $j$ -th guessing game of an agent  $a$  is successful. Otherwise, let  $cs_j^a = 0$ . The definitions of cumulative communicative success  $CS(n)_j^a$  of an agent  $a$ , and of cumulative communicative success  $CCS(n)_j^A$  of a population  $A$ , at game  $j$  for the last  $n$  guessing games are defined in the same way as  $DS(n)_j^a$  and  $CDS(n)_j^A$ , respectively (replace  $DS$  with  $CS$  and  $ds$  with  $cs$ ).

Finally, we also track the number of active words in the lexicon. We call a word  $f \in L$  active if  $[f]_p^L \neq \emptyset$ . The active lexicon of an agent is the set of her active words. It represents the actually used fragment of the entire language of an agent.

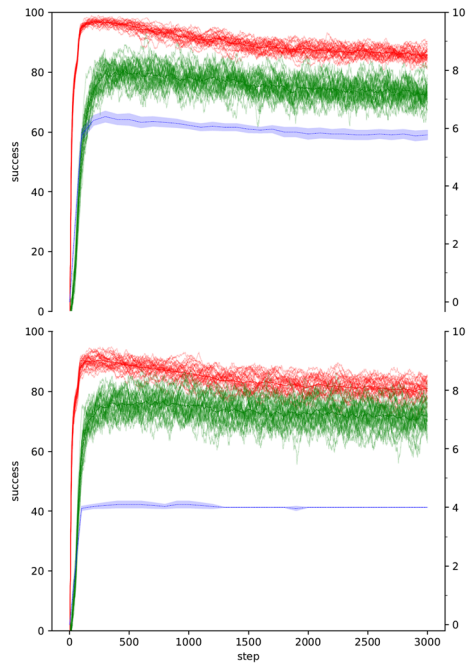
We present validity metrics only for the ANS condition (Fig. 15). All metrics attain higher levels for the condition without ANS. This should not be surprising—turning ANS off leads to more precise number perception and the evolution of more fine-grained terminologies.

We observe that (in the condition with ANS) the discriminative success attains levels mostly above 80%. This shows that agents successfully discriminate the topic from the rest of the context in the vast majority of discrimination games. The communicative success varies between 60% and 90%, with an average between 70% and 80%, which is strictly above the minimal usability success rate of 50%. This shows that the evolving languages are communicatively useful, allowing the interlocutors to successfully communicate about the topic in the majority of guessing games. The languages evolving for numeric-based stimuli have lexicons with 6 active terms on average. The active lexicons for quotient-based stimuli have mean size 4 (see also Figs. 17, 19).

The results support the conclusion that the model defined in Sect. 3 solves the language coordination problem for the domain of quantities. It then may serve as a hypothetical mechanistic theory explaining the transition from the stage of having no quantity terminology to the stage in which such a terminology exists. This allows us to

<sup>11</sup> Note that  $CS$  will never be greater than  $DS$  since successful communication requires successful discrimination.

**Fig. 15** ANS condition: cumulative discriminative success (red), communicative success (green), and mean size of active lexicon (blue, scale on the right) for numeric-based (top) and quotient-based stimuli (bottom). Each red/green line corresponds to a single trial. (Color figure online)



turn our attention to assess the influence of coordination and ANS on the emergence of convexity and monotonicity.

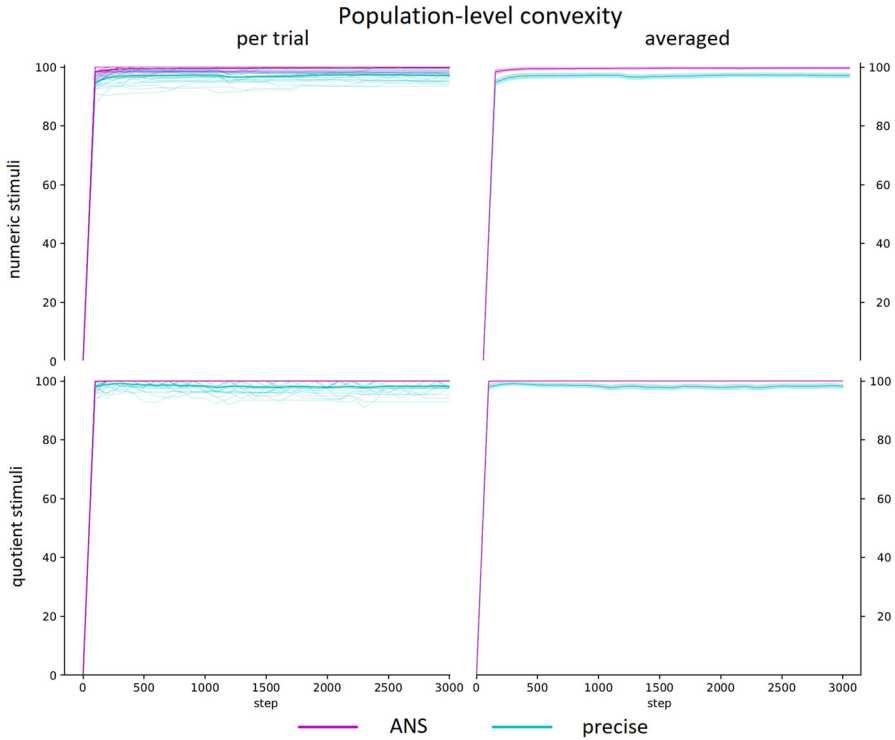
## 5.2 Coordination, ANS and Semantic Universals

We analyse the levels of convexity and monotonicity for languages evolving under four experimental conditions depending on the perceptual system of agents (precise or ANS) and on the type of provided stimuli (numeric or quotient).

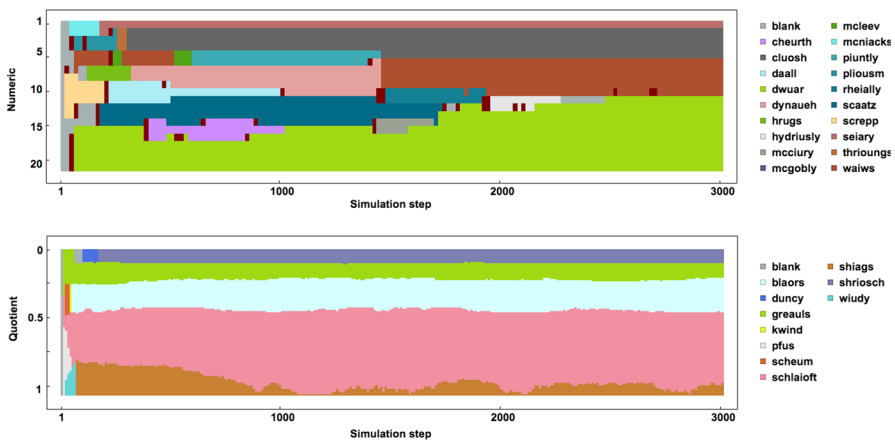
The convexity of a language for a single agent is calculated as the number of active words that have convex pragmatic meanings, divided by the size of the active lexicon. The population-level convexity is the average convexity of languages of all agents in the population.

Figure 16 shows the results. We observe that convexity appears naturally with high frequency in all conditions but also that ANS representation facilitates convexity to a larger extent than the precise representation. The convexity of pragmatic meanings is clearly visible in Fig. 17, where each vertical section represents the active lexicon at a given step. All emerging terminologies are convex—at step 3000 there are no words that contain ‘holes’ in the depicted quantity intervals.

The monotonicity of a language for a single agent is calculated as the number of monotone meanings corresponding to words from the active lexicon, divided by the size of the active lexicon. The population-level monotonicity is given by the average monotonicity of languages of all agents in the population. The plots in Fig. 18 represent population-level monotonicity for all 30 trials.



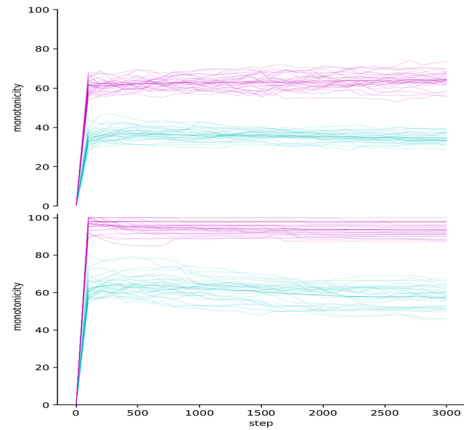
**Fig. 16** Population-level convexity in all four experimental conditions. The y-axes represent percentage scales. Plots on the left track population-level convexity separately for each trial (out of 30). Plots on the right show population-level convexity averaged across trials. Plots on the top and on the bottom correspond to the experimental conditions with numeric-based and quotient-based stimuli, respectively



**Fig. 17** Active lexicon for numeric-based (top) and quotient-based stimuli in the ANS condition for agent 4 (compare with the same agent in Fig. 19)



**Fig. 18** Plots for the numeric-based (top) and quotient-based stimuli (bottom). Condition without ANS is in cyan, while the ANS condition is in magenta



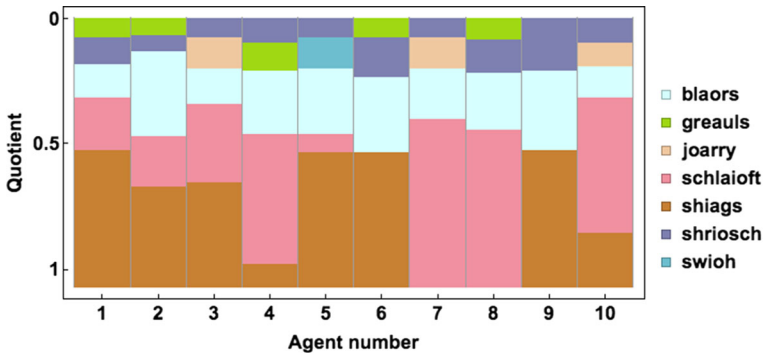
In this case, there is no doubt that ANS contributes significantly to the emergence of monotonicity in language. The difference in monotonicity levels between numeric-based and quotient-based condition is most likely due to the fact that ANS allows for more fine-grained categorization of the number line  $1, 2, \dots, 20$  than the quotient line between 0 and 1.

## 6 Discussion and Conclusions

We presented an agent-based model for the cultural (multi-agent) evolution of quantity expressions (quantifiers), based on the existing guessing-game approach to simulate the emergence of colour expressions (Steels & Belpaeme, 2005). Simulations result in communicatively usable quantity expressions, proving that the model solves the language coordination problem for the domain of quantities and corroborates the robustness of the original model.

Our view of language can be described as liberal, dynamic, and pluralistic. Interestingly, it still relates to the classical discussion in GQT. The use of the classical number triangle representation of quantifiers adds to the existing discussion in three ways. Firstly, we propose that a conceptual space for numerosities could draw inspiration from the number triangle itself, as the latter allows for representing both the distribution of elements in the model, and its (growing) size along two different dimensions. Secondly, we give a new account of convexity through considering and formally defining the property of quantifier continuity. Thirdly, we introduce the new notion of quotient-sensitive quantifiers that contributes to GQT in general. We show how quotient-sensitivity allows a simple view of the regularities of such quantifiers and how various notions of monotonicity contribute to that picture.

To make the connection precise, we showed what kind of generalized quantifiers of type  $\langle 1 \rangle$  correspond to our quantity expressions. We tied the numerical expressions to the properties of EXT and SYMM, and the quotient expressions to the newly proposed quotient-sensitivity. Further, our notions of monotonicity corresponds to  $\text{MON}\uparrow$  and  $\text{MON}\downarrow$ , and convexity corresponds to continuity in the right argument, R-CONT.



**Fig. 19** Active lexicons of all agents from a selected trial at the last step of the simulation for quotient-based stimuli

Clarifying the correspondence with previously known abstract universal properties of generalized quantifiers allows recognizing how a simple coordination procedure may result in evolving a logically rich language.

One of the objectives of this paper ([Question 1](#)) was to see whether certain quantity universals, namely convexity and monotonicity, emerge from the horizontal dynamics of coordination, as implemented in the Steels and Belpaeme language coordination model (Steels & Belpaeme, 2005). What we found is that monotonicity indeed takes off but does not dominate the emergent quantity terminologies. On the other hand, convexity prevails in all conditions. This suggests that the main source of convexity comes from the language coordination model and its dynamics rather than from the details of agents' perceptual system or from the type of stimuli. This is intriguing, as the model itself does not rule out non-convexity (unlike the convexity operator in Pauw and Hilferty, 2012), and it even makes room for it because categories can be formed from distant reactive units or the same word can be linked to distant categories. Anyway, we can conclude that through language coordination, the theory of conceptual spaces gains an additional support from the language evolution perspective.

What sets our approach apart from the original model by Steels et al. (where agents' perception was attuned to colors) is the perception module. One of its parametrizations is based on a well-known psychophysical model of number perception, the so-called approximate number sense (ANS), which assumes that the imprecision of number estimates grows proportionally to the magnitude of the stimulus. Interestingly, languages evolving among agents using only approximate number sense still allow for successful communication. This observation aligns with the idea that ANS can play an important role in the semantics of quantity terms, as some studies have already suggested before (Gordon, 2004).

To get a better understanding of the resulting quantity terminologies, let us take another look at [Fig. 17](#) and compare it to [Fig. 19](#).

Each colour represents a different word, whose semantics spans over a range of stimuli (the y-axis). In [Fig. 17](#), we see active words appearing and disappearing with time. We can also observe that after the initial variation, the terminologies start to regularize structurally, following the ANS-based perception—there are several words

with restrictive fine-grained meaning, for small quantities; for larger ones the words have broader meanings. Figure 19 illustrates the fact that even though the languages of individual agents differ among each other, a satisfactory level of communication is still possible to achieve. This also shows how, in principle, the semantics of quantity expressions can evolve in the absence of precise counting mechanisms.<sup>12</sup>

Our second main objective was to test whether the perceptual constraints of the approximate number sense make the coordinated languages more monotone or more convex (Question 2). We used the precise number sense in the alternative condition—an assumption that prevails in existing explanations of universals.<sup>13</sup> By evolving languages separately under these two conditions, we have found that ANS facilitates both properties, but the positive effect is especially evident in the case of monotonicity.

We can speculate about the causes of the above effect. Here, we mention a few such factors acting at the lower layers of the model and propagating in a bottom-up fashion all the way up to agents' languages. Starting from perception, the boundedness of a linearly ordered scale of numerosities makes certain ANS-based reactive units—especially those closer to the upper boundary—'overspill' outside the scale (see, e.g., Fig. 2). At the level of the discrimination game, the more relaxed, vague characteristics of categories likely allow an easier upward or downward merge with other categories. Since the meaning of a term includes all stimuli that yield positive responses of categories associated with that term, it is natural to expect that many terms would become monotone by virtue of the above-mentioned effects.

The relationship between ANS and convexity is much less clear. As already pointed out, it seems that the high base-levels of convexity stem from other layers of the model. However, convexity is not completely immune to precision. We have observed that ANS facilitates convexity to a larger extent than precise number perception, leading, for the most part, to fully convex terminologies. We can hypothesize that if one is attached to the intuition that cognitive concepts should be convex in the appropriate conceptual space, it is more plausible to assume vague meanings of words.

To relate this work to other studies, it is instructive to realize what are the kinds of pressures that shape languages in our experiments. Clearly, there is a discriminative pressure towards fine-grained categorizations. This pressure is exerted at the level of the discrimination game. There is also a communicative pressure towards mutual comprehensibility, exerted at the level of the guessing game.

One might also argue that some form of simplicity bias is present in the model. For example, one of the hyperparameters ( $\delta$ ) suppresses the expansion of the repertoire of categories of each agent whose current discrimination success is sufficiently high. Also, individual languages do not become unnecessarily complex as associations are only added or strengthened ( $\delta_{inc}$ ) as needed, while useless associations fade away ( $\delta_{inh}$ ,  $\delta_{dec}$ ). There seems to be no other built-in pressures towards simplicity in the model. While the matter may be worthy of deeper analysis, the abovementioned simplicity biases are rather weak when compared to strong forms of built-in complexity

<sup>12</sup> A cursory comparison with the semantics of the Mundurukú quantity terminology (Fig. 2 in Pica et al., 2004) naturally evolved within a community that lacks linguistic means for precise counting, reveals striking similarities.

<sup>13</sup> The work by Pauw and Hilferty (2012) is an exception but it assumes convexity rather than explains it.

minimization used in other studies, be it description length or (approximate) Kolmogorov complexity (Carcassi et al., 2019; Steinert-Threlkeld, 2019; van de Pol et al., 2023). In our model, agents do not monitor the complexity of categories or meanings of words—they rather build them piece-by-piece, indifferent to their complexity.

To conclude, our experiments show that mere language coordination mechanisms lead to (predominantly) convex quantity terminologies. Moreover, our natural perceptual tendencies, as captured by the ANS, lead to more monotone and almost entirely convex languages if only discriminative and communicative pressures are at play—with the reservation that simplicity could be an additional factor resulting from the discrimination and coordination mechanics. However, it would be surprising, but also illuminating, if simple mechanics of the discrimination game and the guessing game significantly restricted the complexity of categories and language and, thus, served as the mechanistic explanation of the simplicity bias.

Other open questions concern understanding which parts of the coordination model are responsible for high levels of convexity in the evolved quantity terminologies (see Fig. 16). This perhaps could be best argued for by alternating various subroutines of the coordination algorithm. Additionally, the behaviour of the model can be tested across different semantic domains. For example, it has not been examined whether colour terminologies from Steels and Belpaeme's original study (Steels & Belpaeme, 2005) are convex, though we hypothesise that this should be the case, at least to some significant degree.

Another set of questions is related to the perception of agents for quotient stimuli. We have tentatively used the most compelling model of ratio perception vaguely suggested in the literature (O'Grady & Xu, 2020). While sufficient for the purposes of the present study (one of which was to examine how presence or absence of approximate quantity estimation affects some characteristics of emergent languages), the model is worth investigating on its own, including confrontation with real data and comparison with other candidate models.

Yet another question is related to the view that, technically, ANS is a combination of vagueness (which can be formalized by assuming a larger standard deviation in Eq. (1), say  $\sigma = 1$  constant across stimuli) and Weber's law. Since our analysis juxtaposes ANS against precise number perception (no vagueness and no Weber's law), we cannot derive any firm conclusions regarding the relative contribution of these two factors to semantic universals. However, we would like to put forth a conjecture that links this relative contribution to the size of the scale of numbers on which agents coordinate their languages. For small scales, vagueness will outperform ANS in terms of monotonicity, because for initial numbers ANS is quite precise and therefore monotonicity of the system of categories will be small (assuming that categories are mostly unimodal). On the other hand, for pure vagueness, monotonicity can be higher because even stimuli in the middle of the scale can generate categories with nonzero responses for stimuli from the ends of the scale. The situation changes, however, if we consider wider scales. Now, Weber's law makes a difference because categories on the right side of the scale tend to over-spill to the right. The wider the scale, the greater the overspill of categories based on ANS (and their monotonicity). On the contrary, driven by the discriminative and communicative pressure, pure vagueness might lead to a more fine-

grained categorization and less monotone languages. Preliminary experiments seem to align with this conjecture but further work is needed to arrive at clear conclusions.

## Supplementary information

The code used to perform simulations and produce some of the plots can be found here: <https://github.com/juszjusz/coordinating-quantifiers>.

**Acknowledgements** All authors were supported by Polish National Science Centre (NCN) OPUS 10 Grant, Reg No.: 2015/19/B/HS1/03292. We would like to thank anonymous reviewers for their comments.

**Funding** Open access funding provided by Technical University of Denmark

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix A Proofs from Sect. 4.4

**Proposition 3** *Let  $Q$  be a quantifier of type  $\langle 1 \rangle$ . If  $Q$  is quotient-sensitive, then  $Q$  is ISOM.*

**Proof** Let us take a type  $\langle 1 \rangle$  quantifier  $Q$  and assume it is quotient-sensitive. Let us now take two models  $M_1 = (U_1, A_1)$  and  $M_2 = (U_2, A_2)$ , such that  $\text{card}(A_1) = \text{card}(A_2) = p$  and  $\text{card}(U_1 - A_1) = \text{card}(U_2 - A_2) = q$ . Then  $\frac{\text{card}(A_1)}{\text{card}(U_1)} = \frac{\text{card}(A_2)}{\text{card}(U_2)} = \frac{p}{q}$ , and so, by quotient-sensitivity,  $Q_{M_1}(A_1) \Leftrightarrow Q_{M_2}(A_2)$ . Hence,  $Q$  is ISOM.  $\square$

**Proposition 4** *Let  $Q$  be a quotient-sensitive quantifier of type  $\langle 1 \rangle$ .  $Q$  is  $\text{MON}\uparrow$  iff  $Q$  is smooth.*

**Proof** ( $\Rightarrow$ ) Take a type  $\langle 1 \rangle$  quotient-sensitive quantifier  $Q$  and assume it is  $\text{MON}\uparrow$ . Let  $Q$  be the quanti-relation of  $Q$ . First observe that, since  $Q$  is quotient-sensitive, we have that  $(*)$  if  $Q(k, n)$ , then  $Q(pk, pn)$  for any positive integer  $p$ . This follows from the fact that  $\frac{n}{k+n} = \frac{pn}{p(k+n)} = \frac{pn}{pk+pn}$ .

We need to show that:

- ( $\nearrow$ ) if  $Q(k, n)$  and  $k \neq 0$ , then  $Q(k - 1, n)$
- ( $\searrow$ ) if  $Q(k, n)$  then  $Q(k, n + 1)$

For ( $\nearrow$ ) assume that  $Q(k, n)$  with  $k \neq 0$ . Then, from  $(*)$  we get that  $Q((k + n - 1)k, (k + n - 1)n)$ , which can be rewritten into  $Q(k^2 + nk - k, n^2 + nk - n)$ . Since  $Q$  is  $\text{MON}\uparrow$ , i.e.,  $Q$  is  $(\rightarrow)$ , we have that  $Q(k^2 + nk - k - n, n^2 + nk)$  (by ‘moving’

$n$  elements from the first to the second argument). Now, since

$$\frac{n^2 + kn}{(k^2 + nk - k - n) + (n^2 + nk)} = \frac{(k + n)n}{(k + n)(k + n - 1)} = \frac{n}{n + k - 1},$$

by the fact that  $Q$  is quotient sensitive, and  $k \neq 0$ , we get that  $Q(k - 1, n)$ .

An analogous argument can be constructed for  $(\searrow)$ .

( $\Leftrightarrow$ ) Take a type  $\langle 1 \rangle$  quotient-sensitive quantifier  $Q$  and assume it is smooth. Let  $Q$  be the quanti-relation of  $Q$ . Assume that  $Q(k, n)$ . Then, by smoothness, we know that  $Q$  is  $(\nearrow)$ , and so  $Q(k - 1, n)$ . Applying smoothness again,  $(\searrow)$ , we get that  $Q(k - 1, n + 1)$ . We conclude that  $Q$  is  $(\rightarrow)$ , and so  $Q$  is  $\text{MON}\uparrow$ .  $\square$

## References

- Bach, E., Jelinek, E., Kratzer, A., & Partee, B. H. (Eds.). (1995). *Quantification in natural languages. Studies in Linguistics and Philosophy* (Vol. 54). Dordrecht: Springer. <https://doi.org/10.1007/978-94-017-2817-1>
- Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4(2), 159–219.
- Cantlon, J. F., Brannon, E. M., Carter, E. J., & Pelphrey, K. A. (2006). Functional imaging of numerical processing in adults and 4-y-old children. *PLOS Biology*. <https://doi.org/10.1371/journal.pbio.0040125>
- Carcassi, F. (2020) Cultural evolution of scalar categorization: How cognition and communication affect the structure of categories on scalar conceptual domains. PhD thesis, The University of Edinburgh.
- Carcassi, F., Schouwstra, M., & Kirby, S. (2019) The evolution of adjectival monotonicity. In *Proceedings of Sinn und Bedeutung* (Vol. 23, pp. 219–230).
- Carcassi, F., Steinert-Threlkeld, S., & Szymanik, J. (2019). The emergence of monotone quantifiers via iterated learning. In A. K. Goel, C. M. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st annual meeting of the cognitive science society* (pp. 190–196). Montreal: Cognitive Science Society.
- Chemla, E., Buccola, B., & Dautriche, I. (2019). Connecting content and logical words. *Journal of Semantics*, 36(3), 531–547. <https://doi.org/10.1093/jos/ffz001>
- Cheyette, S. J., & Piantadosi, S. T. (2020). A unified account of numerosity perception. *Nature Human Behaviour*, 4(12), 1265–1272. <https://doi.org/10.1038/s41562-020-00946-0>
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Chicago: MIT Press.
- Christiansen, M. H., & Chater, N. (2016). *Creating language: Integrating evolution, acquisition, and processing*. Chicago: MIT Press.
- Croft, W. (1990). *Typology and universals*. Cambridge: Cambridge University Press.
- Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York, USA: Oxford University Press.
- Díaz-Francés, E., & Rubio, F. J. (2013). On the existence of a normal approximation to the distribution of the ratio of two independent normal random variables. *Statistical Papers*, 54(2), 309–323. <https://doi.org/10.1007/s00362-012-0429-2>
- Fechner, G. (1966). *Elements of psychophysics*. Vol. I. New York, Holt, Rinehart and Winston. First published in 1860.
- Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, 8(7), 307–314. <https://doi.org/10.1016/j.tics.2004.05.002>
- Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Chicago, USA: MIT Press.
- Gärdenfors, P. (2014). *The geometry of meaning: Semantics based on conceptual spaces*. Chicago, USA: MIT Press.
- Gordon, P. (2004). Numerical cognition without words: Evidence from Amazonia. *Science*, 306(5695), 496–499. <https://doi.org/10.1126/science.1094492>

- Greenberg, J. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. Greenberg (Ed.), *Universals of language* (pp. 73–113). Cambridge, MA.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1). New York, USA: Wiley.
- Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the “number sense”: The approximate number System in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, *44*(5), 1457.
- Halmos, P. R. (1976). *Measure theory*. New York, USA: Graduate Texts in Mathematics. Springer.
- Haskins, G.: Gibberish. <https://pypi.org/project/gibberish/>
- Hinkley, D. V. (1969). On the ratio of two correlated normal random variables. *Biometrika*, *56*(3), 635–639.
- Hyman, L. M. (2008). Universals in phonology. *The Linguistic Review*, *25*(1–2), 83.
- Jäger, G. (2010). Natural color categories are convex sets. In M. Aloni, H. Bastiaanse, T. de Jager, & K. Schulz (Eds.), *Logic, language and meaning* (pp. 11–20). Berlin, Heidelberg: Springer.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *Statistical learning* (pp. 15–57). New York, NY: Springer. [https://doi.org/10.1007/978-1-4614-7138-7\\_2](https://doi.org/10.1007/978-1-4614-7138-7_2)
- Keenan, E., & Paperno, D. (Eds.). (2012). *Handbook of quantifiers in natural language. Studies in Linguistics and Philosophy* (Vol. 90). Dordrecht, The Netherlands: Springer.
- Keenan, E. L., & Stavi, J. (1986). A semantic characterization of natural language determiners. *Linguistics and Philosophy*, *9*(3), 253–326.
- Lindsey, D. T., & Brown, A. M. (2009). World Color Survey color naming reveals universal motifs and their within-language diversity. *Proceedings of the National Academy of Sciences*, *106*(47), 19785–19790. <https://doi.org/10.1073/pnas.0910981106>
- Lindström, P. (1966). First-order predicate logic with generalized quantifiers. *Theoria*, *32*, 186–95.
- Matthewson, L. (2008). *Quantification: A cross-linguistic perspective. North-Holland Linguistic series: Linguistic variations series* (Vol. 64). Leiden, The Netherlands: Brill.
- Menninger, K. (1969). Number words and number symbols. A cultural history of numbers. Cambridge, MA: M.I.T. Press. Translated from the revised German edition (Göttingen, 1958) by Paul Broneer.
- Mostowski, A. (1957). On a generalization of quantifiers. *Fundamenta Mathematicae*, *44*(1), 12–36.
- O’Grady, S., Griffiths, T.L., & Xu, F. (2016) Do simple probability judgments rely on integer approximation? In *Proceedings of the 40th annual conference of the cognitive science society*, Philadelphia, pp. 1008–1013.
- O’Grady, S., & Xu, F. (2020). The development of nonsymbolic probability judgments in children. *Child Development*, *91*(3), 784–798. <https://doi.org/10.1111/cdev.13222>
- Pauw, S., & Hilferty, J. (2012). The emergence of quantifiers. In *Experiments in cultural language evolution* (Vol. 3, pp. 277–304). Amsterdam, The Netherlands: John Benjamins Publishing Company.
- Peters, S., & Westerståhl, D. (2006). *Quantifiers in language and logic*. Oxford: Oxford University Press.
- Pham-Gia, T., Turkkan, N., & Marchand, E. (2006). Density of the ratio of two normal random variables and applications. *Communications in Statistics: Theory and Methods*, *35*(9), 1569–1591.
- Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian Indigene Group. *Science*, *306*(5695), 499–503. <https://doi.org/10.1126/science.1102085>
- Steels, L. (1997) Constructing and sharing perceptual distinctions. In: van Someren, M., Widmer, G. (eds.) *Machine learning: ECML-97: 9th European conference on machine learning prague, Czech Republic, April 23–25, 1997 Proceedings* (pp. 4–13). Berlin, Heidelberg: Springer. [https://doi.org/10.1007/3-540-62858-4\\_68](https://doi.org/10.1007/3-540-62858-4_68)
- Steels, L. (2012). *Experiments in cultural language evolution. Advances in interaction studies* (Vol. 3). Amsterdam, The Netherlands: John Benjamins Publishing.
- Steels, L., & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Sciences*, *28*(4), 469–489.
- Steinert-Threlkeld, S. (2019). Quantifiers in natural language optimize the simplicity/informativeness trade-off. In J. J. Schlöder, D. McHugh, & F. Roelofsen (Eds.), *Proceedings of the 22nd Amsterdam Colloquium* (pp. 513–522).
- Steinert-Threlkeld, S., & Szymanik, J. (2019). Learnability and Semantic Universals. *Semantics & Pragmatics*. <https://doi.org/10.3765/sp.12.4>
- Szymanik, J., & Zajenkowski, M. (2010). Comprehension of simple quantifiers: Empirical evaluation of a computational model. *Cognitive Science*, *34*(3), 521–532. <https://doi.org/10.1111/j.1551-6709.2009.01078.x>



- Trinh, T. (2019). Exhaustification and contextual restriction. *Frontiers in Communication*. <https://doi.org/10.3389/fcomm.2019.00047>
- van Benthem, J. (1986). *Essays in logical semantics*. Dordrecht: D. Reidel.
- van de Pol, I., Steinert-Threlkeld, S., & Szymanik, J. (2019). Complexity and learnability in the explanation of semantic universals of quantifiers. In: Proceedings of the 41st annual meeting of the cognitive science society.
- van de Pol, I., Lodder, P., van Maanen, L., Steinert-Threlkeld, S., & Szymanik, J. (2023). Quantifiers satisfying semantic universals have shorter minimal description length. *Cognition*, 232, 105150. <https://doi.org/10.1016/j.cognition.2022.105150>
- van Tiel, B., Franke, M., & Sauerland, U. (2021). Probabilistic pragmatics explains gradience and focality in natural language quantification. *Proceedings of the National Academy of Sciences*, 118(9), 2005453118. <https://doi.org/10.1073/pnas.2005453118>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.