Cognition 261 (2025) 106129

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Full length article

Semantic extension in a novel communication system is facilitated by salient shared associations

Kenny Smith [a,*], Josephine Bowerman [b], Andrew D.M. Smith [c]

[a] Centre for Language Evolution, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, 3 Charles Street, Edinburgh, EH8 9AD, United Kingdom
[b] Unaffiliated
[c] Literature and Languages, Faculty of Arts and Humanities, University of Stirling, Stirling, FK9 4LA, United Kingdom

ARTICLE INFO

Keywords:
Language change
Semantic extension
Experimental semiotics

ABSTRACT

Creative processes of semantic extension play a key role in language change, grammaticalisation, and (by hypothesis) the early origins and evolution of language. In this paper we report two dyadic interaction experiments studying the semantic extension of novel labels in controlled circumstances. We find that participants can use salient and shared associations in their perceptual environment (between colours and shapes) to bootstrap a communication system, and can then extend those labels figuratively, to convey both concrete and abstract targets, by exploiting shared understandings such as colours associated stereotypically with specific objects and emotions. By manipulating the presence of reliable statistical associations between colour and shape early in this process we show that such shared associations facilitate both an initial semantic extension and subsequent chaining of extensions; we also find that extensions relying on less certain grounding (e.g. between colours and emotions) lead to greater variability in how extensions are made. Our method can be used to test the creative processes of semantic extension under controlled conditions, and provides experimental purchase on the relationship between association and extension which have only previously been studied through correlational means.

1. Introduction

Human language provides enormous flexibility, allowing its users to devise innovative ways of expressing new ideas and meanings which can nonetheless be understood by others who have never previously encountered them. Such creativity is a result of the ostensive-inferential nature of communication, in which a signal is not part of a fixed code which explicitly specifies a particular meaning, but rather where a signal is used as a piece of evidence to prompt and assist the addressee in reconstructing the initiator's intended message (Scott-Phillips, 2014). This requires interlocutors to perform sophisticated reasoning about each other's intentions, based on the common ground, or mutual knowledge, that they assume they share (Clark, 1996), including not only recognition of each other's communicative intentions and what is relevant in the communicative episode (Sperber & Wilson, 1995; Tomasello et al., 2005), but also an understanding of their shared conventions, both linguistic and non-linguistic. This awareness of common ground enables an initiator to anticipate an addressee's likely inferences in a given context, and thus to provide the most appropriate signals in order to achieve their immediate communicative goals. Crucially, these inferences can only ever be approximate reconstructions of the initiator's intended message, as there are always multiple possible interpretations for any utterance. It is just this lack of precision, however, which provides language with the flexibility to extend signals creatively in order to evoke new, non-conventional meanings.

There are numerous potential ways in which ad-hoc semantic extensions can be motivated (see e.g. Brochhagen et al., 2023, for a recent quantitative exploration), but two of the most important and pervasive are metonymy and metaphor (Dancygier & Sweetser, 2014; Lakoff & Johnson, 2003). In both cases, connections are made between an expression's original meaning and the innovative extended meaning, but while metonymy derives ultimately from the repeated co-incidence of two objects or events in our experience, metaphor derives from perceived resemblances which enable mappings to be made between the entities. Frequently, for instance, metonymic relationships can be found between an object and its most contextually relevant part or attribute (e.g. the meaning of *wheels* being extended to 'car'), or between a location and an important organisation or salient activity which takes place there (e.g. the meaning of the Belgian capital *Brussels* being extended to the 'European Union', whose principal institutions are

* Corresponding author.
E-mail address: kenny.smith@ed.ac.uk (K. Smith).

https://doi.org/10.1016/j.cognition.2025.106129
Received 26 March 2024; Received in revised form 14 March 2025; Accepted 20 March 2025

located there). Metaphorical connections between domains can often be made through physical resemblance — such as, for instance, the extension of the meaning of *mouth* from an entrance into the human body to an entrance into e.g. a river, cave, or bottle (Klepousniotou & Baum, 2007; Klepousniotou et al., 2008) — but also through more abstract connections through the representation of one activity in terms of another, for example the extension of the meanings of words like *chew*, *swallow* and *digest* from describing the eating of food to the mental consideration of ideas (Kövecses, 2010). Metaphor and metonymy cannot always clearly be distinguished, however, particularly with respect to so-called correlation metaphors, which likely derive from earlier metonymic representations where one element of the metonymy has been generalised to a concept in a different domain and then linked metaphorically to the other element. The source of the metaphor SAD IS DOWN, for example, is arguably the metonymic relationship DOWNWARD BODILY POSTURE FOR SADNESS, where an effect is used to represent its cause, followed by the generalisation of the specific concept 'downward bodily posture when sad' to the more general concept 'down' in the domain of space (Kövecses, 2020).

In each case, the successful interpretation of such creative semantic extensions depends on the addressee's focus on what is most relevant in the immediate communicative context, which in turn determines the aspects of the signal's conventional meaning which should be ignored, and the innovative semantic elements which should be brought to the fore (Smith & Hoefler, 2017). Although any single semantic innovation is inherently ephemeral and concerned only with short-term communicative goals, some successful inferences will inevitably occur more frequently than others, due to natural connections between existing and innovative meanings. If a particular usage recurs sufficiently frequently in similar contexts, then the inferential processing required to interpret the innovative meaning may become increasingly routinised; this strengthens the association between the form and the novel meaning, until it may turn into a linguistic convention in its own right, perhaps eventually becoming sufficiently entrenched to be available even without the scaffolding of the original communicative context (Kuteva, 2008; Traugott & Dasher, 2001). Once conventionalised, the new meaning itself becomes readily available to serve as the basis for further innovative semantic extensions. This chaining of repeated extension and conventionalisation (Ramiro et al., 2018) leads, through the ratchet effect (Tomasello, 1999), to increases in the complexity of meanings which can be communicated, as formerly inaccessible meanings are reached through a sequence of stepping stones (Hoefler, 2009).

A similar process can be seen in the complexification of language through grammaticalisation, where lexical, concept-encoding items turn into functional items with more abstract, grammatical meanings (Hopper & Traugott, 2003), and which has been proposed as the main mechanism generating the syntactic complexity characteristic of natural languages (Deutscher, 2005; Heine & Kuteva, 2007; Smith & Hoefler, 2017). Semantic extension through both metonymy and metaphor is widely recognised in historical linguistics as a key source of the relevant linguistic innovations which afford the inference of more abstract meanings in addition to their existing meanings, and eventually lead to grammaticalisation (see e.g. Bybee et al., 1994; Givón, 1979; Haspelmath, 1998; Heine, 1997; Heine et al., 1991; Traugott & Dasher, 2001). For instance, markers of temporal sequences frequently develop into more grammaticalised causal markers through metonymy (e.g. English *since* 'because of' < Old English *siþþan* 'from the time that') because causes typically directly precede their effects, and so the same construction can plausibly and naturally be interpreted either sequentially or causally (Traugott & Dasher, 2001; Traugott & König, 1991). Many metaphorical extensions, too, lead to grammaticalised constructions; one particularly fruitful class of extension is the use of distinctive parts of the human body to express similar parts of other objects, such as the leg of a table or the eye of a needle (Kuteva et al., 2019). A metaphorical meaning like this can then itself often be extended to refer to the area next to the part of the object, and from there to a more generalised location, for example English *in front of* < *front* 'area by the main side of a building' < French *front* 'forehead' (Svorou, 1994).

The successful expression and interpretation of novel meanings, therefore, is a crucial mechanism in the process through which linguistic complexity emerges and is maintained. While this process is often studied from a historical perspective, inferring semantic change in historical corpora and making inferences about the underlying metonymic and metaphorical processes involved (see references above), or from a typological perspective, identifying recurring cross-linguistic pathways of semantic change (e.g. Kuteva et al., 2019), semantic extension is ultimately grounded in communicative interaction between individuals, as they strive to convey their intended meaning using the linguistic tools at their disposal, and the subsequent learning and adoption of those innovations by the wider community as they conventionalise (or fail to). Over recent decades, controlled laboratory experiments with artificial languages have been productively used to investigate the role of learning and communicative use in shaping key properties of language which have previously only been studied from a historical or typological perspective, providing a means to directly test mechanisms and constraints inferred from historical or typological data. These methods have been applied to a wide range of phenomena, including the development of fundamental structural features shared by all languages (such as compositional morphosyntax, Kirby et al., 2015; Raviv et al., 2019a, combinatorial phonology, Verhoef, 2012; Verhoef et al., 2014, and coordinated semantic conventions, Galantucci, 2009; Scott-Phillips et al., 2009), typological universals in syntax and phonology (e.g. Culbertson & Adger, 2014; Culbertson et al., 2012; Martin & White, 2019; White, 2014; see Culbertson, 2012 for review), or recurring tendencies in language change such as the effects of frequency on preservation of irregularity (Smith et al., 2023) or the tendency for optional elements to become obligatory (Fehér et al., 2019).

Of particular relevance here are experimental studies from evolutionary linguistics, based around communicative interaction (e.g. Galantucci, 2009; Roberts et al., 2015; Scott-Phillips et al., 2009; Verhoef et al., 2016), which study semantic extension in controlled circumstances. In this latter type of study, known as experimental semiotics, participants develop new shared communication systems from scratch over the course of repeated interactions, often using novel signalling channels, e.g. choosing colour shades from a continuous spectrum (Roberts & Clark, 2020), or with finger movements on a vertical bar mapped to movement of a dot on-screen (Verhoef et al., 2016; see also Verhoef et al., 2024). Coordinating the allocation of meanings to novel signals is a difficult task in which participants need to anticipate the inferences likely to be made by their interlocutor in response to a signal. For instance, Verhoef et al. (2016) study the emergence of conventions for conveying temporal meanings (e.g. *yesterday*, *tomorrow*, *last year*, *next year*) in a novel communication medium where visual signals are generated by finger movements (one participant sees a moving dot replaying the vertical component of the finger movement of their partner). Verhoef et al. find that their participants reliably exploit a shared association between temporal duration and spatial extent to convey differences of temporal duration, i.e. longer temporal durations are conveyed using signals with a larger spatial extent, metaphorically extending spatial signals into the temporal domain.

In this paper we adapt these methods to study semantic extension in controlled circumstances. We test the conditions under which labels for specific abstract meanings can be established and maintained by interlocutors, to shed light on the hypothesised initial stages of both the origin of language and its subsequent complexification. Participants played communication games in pairs, choosing from a limited array of geometric shapes to use as labels to communicate with their partner about distinct shapes, colours, objects and emotions. We use this paradigm in two experiments to test whether shape labels can be extended to represent colours which they co-occurred with in the initial stages of the experiment; we manipulate the consistency of shape–colour associations to test whether, as is widely assumed, the availability of salient and shared metonymic associations affects participants' ability to productively use these extensions in communication. In later stages of the experiments we test whether participants can chain such semantic extensions, using the shape labels not just to represent the colour meanings, but to extend these meanings figuratively in two ways, based on different real-world contingencies: (1) further metonymic extension to refer to objects for which the relevant colour is a very salient attribute in the real world (e.g. because a stereotypically relevant feature of pigs is their pink colour, can participants use their established label which means 'pink' to refer to a pig?); (2) further metaphorical extension through a longer and more indirect chain of reasoning to use shapes to convey emotions (e.g. because people are frequently happy when the sun shines, and the sun is yellow, can participants use the label which has come to mean 'yellow' to refer to happiness?)[1] We find that initial semantic extension is facilitated by salient and shared metonymic associations, and that the strength of grounding of those initial extensions results in more robust extension behaviour in subsequent chained extensions, providing support to the hypothesis that exploiting common ground is key to the creatively flexible conventions which underpin language.

2. Experiment 1

Participants played a communication game with a partner, taking turns sending and interpreting simple visual labels (shapes). Participants worked through six blocks of increasing difficulty, with later blocks requiring them to extend their established labelling conventions to convey additional, more abstract meaning distinctions (colours, objects and emotions). In the early part of the experiment we manipulated the presence of reliable associations between shapes and colours, and thus manipulated shared grounds for the extension of shape terms to refer to colour in later blocks.

2.1. Methods

2.1.1. Participants

268 English-speaking adults were recruited through the online crowdsourcing platform Prolific. We obtained complete data from 107 pairs (214 participants), 54 pairs in the Fixed Associations condition and 53 in the Random Associations condition. We used Prolific's filters to restrict the experiment to people who self-identified English as their first language, and we did not use any geographical restrictions; for participants for who we have geographical information (via summary demographic information provided automatically by Prolific), 66% gave the UK as country of birth (12% gave South Africa, 3% gave USA, 2% gave Australia, Canada, Ireland, New Zealand and Zimbabwe, all other countries of birth were represented at 1% or less each), and 70% gave the UK as their country of residence (15% South Africa, 3% USA, 2% Australia, Canada and Ireland, other countries of residence were represented at 1% or less each). Participants who completed the experiment (median completion time 41 min) were paid £7; partial payment was made to participants who were unable to complete the experiment (e.g. due to not being paired with a partner).

2.1.2. Materials

We selected a set of simple geometric shapes to serve as the labels in a graphical communication game, and in some phases of the experiment as the referents to which those labels referred (see below). Each pair of participants was allocated 4 shapes selected at random from a set of 7 options (square, circle, diamond, star, cross, pentagon, hexagon). These shapes were chosen to be easily distinguishable from one another and to avoid resemblance to any of the target pictures that they would be used to communicate about at later blocks (e.g. we ruled out using a triangle because of its visual similarity to the volcano picture in the Objects block; see below). We used shapes as labels (rather than e.g. novel words) because (i) we expected them to have relatively weak pre-existing associations with colours, objects and emotions (not entirely correctly, as it turns out), allowing us more control in testing the availability of metonymic and metaphorical associations to processes of semantic extension, and (ii) we (correctly) expected shapes to have iconic affordances in the first blocks of the experiment, allowing participants to quickly grasp the communicative task without additional training.

Some blocks of the experiment involve shapes or coloured splats appearing in one of four different colours: red, yellow, pink, and grey. These colours have been shown to have strong, stable associations with specific emotions, both for native English speakers and across cultures (Jonauskaite et al., 2020, 2019): anger (red), happiness (yellow), love (pink), and sadness (grey). This led us to hypothesise that these colour-emotion mappings would be available to participants as grounds for metaphorically-motivated semantic extensions in the Emotions block of the experiment (see below). Further, we selected a set of simple black-and-white line drawings of four different objects that we judged to be associated with each of the four colours by virtue of the colour being a stereotypically relevant feature of the object: volcano (red), banana (yellow), pig (pink), and city (grey).[2] Lastly, we chose four photographs of people representing the emotions that Jonauskaite et al. (2020, 2019) argue to be associated with our chosen colours. These pictures were obtained by google image search on the relevant emotion terms.

2.1.3. Procedure

The experiment was coded in JavaScript using jsPsych (de Leeuw, 2015) and featured real-time interaction between crowdsourced participants, achieved via WebSockets and a python server coordinating pairs of participants (based on code from https://kennysmithed.github.io/oels2024/; the full code for Experiment 1 and Experiment 2 is available at https://github.com/kennysmithed/SemanticExtension).

Participants were informed that they would be playing a communication game with a partner, which would involve sending and interpreting simple visual labels. Before being matched with a partner, each participant completed a screening task to check for colour-blindness, and worked through a short warm-up that involved communicating with the computer, taking turns to play as Sender and Receiver with a trial procedure similar to that in the main task but with different target meanings. Participants failing a colour blindness trial or making more than one mistake on the (trivial) warm-up were ejected from the experiment. Participants were asked not to take written notes; 5 participants admitted to taking written notes as part of the post-experiment debrief and were excluded from the analysis, together with their partners, but paid in full.[3]

[1] This type of extension could be regarded either as a complex chain of metonymy, or as a correlational metaphor derived from earlier metonymies. For simplicity, in the experiments below, these two semantic extensions are described as metonymic and metaphorical extensions respectively.

[2] These drawings were used under license from https://thenounproject.com. Banana created by Icon Producer; city created by Made x Made; pig created by Laymik; volcano created by Nociconist.

[3] Participants received two instructions about taking written notes: an initial instruction ("*Please do not take written notes!* In this experiment we are interested in what your brain can do, not what your brain plus a notebook can do, so please don't write anything down. Just do your best — we are interested

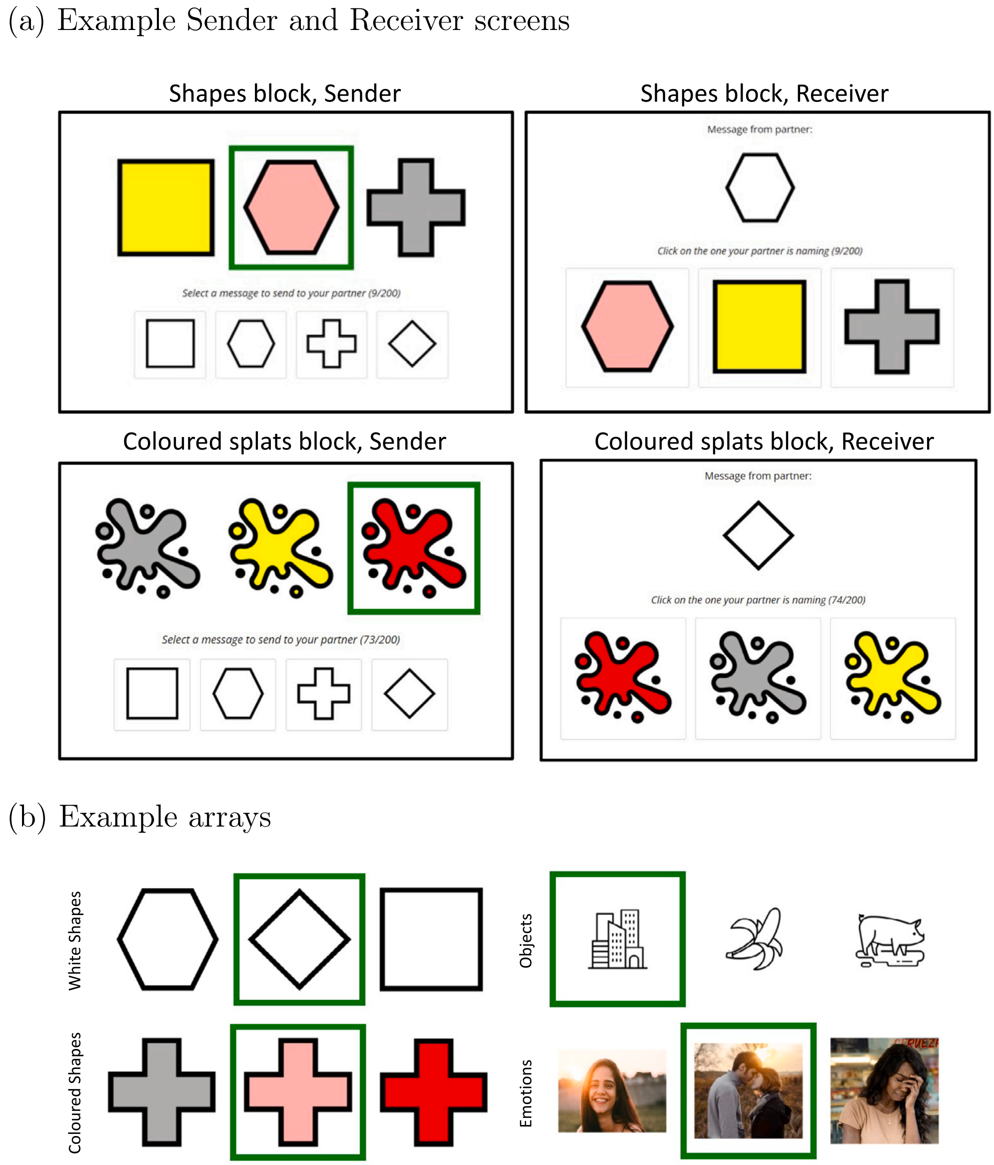

Fig. 1. (a): Example Sender and Receiver screens from the Shapes (upper) and Coloured Splats (lower) blocks; the Sender selects a label (a geometric shape) to convey the target picture (highlighted with a green box) to the Receiver, who selects a picture based on that label. (b): Example arrays in the other four blocks (White Shapes; Coloured Shapes; Objects; Emotions), as seen by the Sender, with the target highlighted. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

After consent and instructions, participants entered a waiting room and were paired with another participant for the communication game proper.[4] Paired participants worked through six blocks of increasing difficulty, taking turns as *Sender* and *Receiver* and switching roles after every trial. On any trial (see Fig. 1) the Sender saw an array of three pictures, with the target picture highlighted with a green box. The Sender's task was to select a label (a geometric shape from their label inventory at the bottom of the screen) to label the target picture for the Receiver. The Receiver then saw the Sender's selected label, plus the same array as seen by the Sender but with left–right order randomised and no highlighting of the target; the Receiver's task was to click on the picture being labelled by the Sender. Both Sender and Receiver were awarded a point for every correct Receiver choice, and both Sender and Receiver saw as feedback the Sender's target picture, the label used, and the Receiver's selection. Success on this communicative task reflects participants' ability to establish a shared system of form-meaning mappings. The *White Shapes* block (the first block of the experiment) was

in places where communication breaks down as well as where it succeeds, and we don't expect perfection!") then a reminder immediately prior to interaction ("*Remember*, we ask you not to take written notes during the experiment — we are interested in what your brain can do, not what your brain plus a notebook can do, so just do your best".) After completing the experiment, as part of the debrief procedure we checked for compliance with this instruction: "Did you write stuff down or take notes during the task? Please be honest — it won't affect your payment, we promise, and if you tell us now we can correct for this in our analysis without affecting the validity of our experiment". and participants selected from two options, either "No I did not make notes" or "OK I confess, I did make notes!".

[4] The maximum waiting time in the waiting room was 5 min; participants who were not paired in this time exited the experiment and received partial compensation. In practice very few participants ended up not being paired, since we found that newly-posted experiment slots on Prolific tended to be taken up rapidly, resulting in a flurry of participants arriving in the waiting room at approximately the same time and being paired.

intended to familiarise participants with communicating using shapes as labels: all the pictures in the Sender and Receiver's arrays were of white shapes (see Fig. 1), meaning that the sender could trivially encode the target picture by using the corresponding shape as the label. This block featured 8 trials (each participant acting as Sender for each shape once).

In the *Shapes* block (block 2) the shapes in the array (but not the labels) appeared in colour (grey, red, pink or yellow); however, all items in the array also differed in shape, meaning that encoding the shape of the target was again trivial and sufficient for successful communication. We used the Shapes block to establish two conditions. In the *Fixed Associations* condition, each shape was assigned a colour and always appeared in the Sender or Receiver's array in that colour (e.g. the hexagon was always pink), establishing an association between shapes and colours, and thus shared grounds for the extension of shape labels to refer to colour in later blocks (each pair was allocated a consistent randomly-generated shape–colour assignment). In the *Random Associations* condition, the colour in which a shape appeared in the array was randomly generated on each trial (and was the same for both Sender and Receiver); colours for each object were selected at random with uniform probability from among the options grey, red, pink or yellow, with the constraint that no two shapes in the array could have the same colour.[5] The Shapes block featured 64 trials, each participant acting as Sender for each target shape 8 times.[6]

In the *Coloured Splats* block (block 3), the array on each trial featured three identical splats differing only in colour (see Fig. 1). Participants therefore needed to establish shared mappings between the target colours and the shape labels. We expected that the colour–shape associations established during the Shapes block would make this straightforward in the Fixed Associations condition, allowing the participants to extend the meaning of those labels to the associated colour (e.g. using the hexagon label to convey the colour 'pink' if hexagons were pink in the Shapes block); we expected this would be substantially harder for participants in the Random Associations condition, given the lack of such prominent and shared associations (i.e. communicative success would be lower). In the *Coloured Shapes* block (block 4), the array featured three of the same shape (e.g. three crosses) in different colours. To communicate the target shape (e.g. the pink cross), participants again had to draw on a shape–colour mapping to signal the colour of the target, but in this case also over-riding the label's inherent, iconic meaning (e.g. using a hexagon label to refer to the pink cross, because hexagons were pink in the Shape block, rather than a cross label which would not uniquely identify the target).

The Coloured Splats and Coloured Shapes blocks therefore provided a first test of whether participants could successfully extend the meaning of a shape label from its inherent meaning (e.g. the star label means star) to encompass colour (e.g. the star label means red). The Coloured Splats and Coloured Shapes blocks each consisted of 32 trials (each participant acting as Sender for each colour 4 times in each block).

The final blocks tested whether participants were able to make further extensions drawing on pre-existing associations of the extended colour meaning. In the *Objects* block (block 5), the array consisted of three black-and-white line drawings of physical objects/entities (on each trial, 3 selected from: volcano, banana, pig, city). Participants had to select a shape label to communicate the target to their partner, e.g. by using the shape–colour associations established in the previous blocks, and extending them based on metonymic associations between an object and its salient stereotypical colour in the real world (e.g. the star label means 'star' and/or 'red'; a volcano is stereotypically red; therefore the star label can be used creatively to mean 'volcano'). Finally, in the *Emotions* block (block 6), the array consisted of three photographs representing different abstract emotions (on each trial, 3 selected from: anger, happiness, love, sadness). Participants again had to select a shape label to communicate the target to their partner, e.g. using their established shape–colour associations, but this time the crucial semantic extensions were based on correlational metaphors based on associations between colours and emotions (e.g. the star label means 'star' and/or 'red'; anger is associated with the colour red; therefore the star label can be used creatively to refer to 'anger'). These final two blocks were each 32 trials long (each participant acting as Sender for each object/emotion 4 times in each block).

2.2. Results

2.2.1. Communicative success

As expected, communicative success (the proportion of trials where the Receiver clicked on the correct response, i.e. the target item the Sender was attempting to convey with their label) in the first two blocks of the experiment (White Shapes, Shapes) was uniformly high, with average accuracy of over 99% in both conditions. Fig. 2 shows communicative success on the subsequent blocks of the experiment. Participants were quite successful in the task in both Fixed Associations and Random Associations conditions, and many pairs were able to reach high levels of accuracy in using shapes to communicate about colour (in the Coloured Splats and Coloured Shapes blocks), could decouple the basic (shape) and extended (colour) meaning of the labels in the Coloured Shapes block (e.g. using the star label to refer to a red square because star can also mean 'red'), and were able to further extend those colour–shape associations creatively to communicate about objects and emotions. As expected, participants in the Random Associations condition appear to be at a disadvantage in this semantic extension process, due to the lack of any consistent shared statistical associations between colour and shape in their Shapes block; this is particularly noticeable in the Coloured Splats block, the first block where participants have to extend the meaning of the labels to communicate about colour.

We used mixed effects logistic regression to analyse the binary outcome of communicative success on each trial (success/failure), with condition, block, and their interaction as fixed effects.[7] A model with block treatment-coded (with Coloured Splats as the reference level) and condition (Fixed or Random Associations) sum-coded shows that the conditions differed at the Coloured Splats block ($b=0.32$, $SE=0.12$, $p=.009$), with participants in the Fixed Associations condition communicating more successfully. A model with block coded using successive differences shows that performance on the Coloured Shapes block was higher than in the Coloured Splats block ($b=0.26$, $SE=0.09$, $p=.004$) and was higher in the Objects block than the Coloured Shapes block ($b=0.56$, $SE=0.13$, $p<.001$), but there was no increase in performance from Objects to Emotions blocks ($b=0.16$, $SE=0.12$, $p=.182$). While the difference between conditions is numerically largest in the Coloured Splats block and declines across subsequent blocks, neither model showed a significant interaction between block and condition (e.g. in the treatment-coded model, lowest $p=.187$). These results therefore

[5] An interesting avenue for future research, which we do not pursue here, would be to explore levels of association that lie between these two extremes, and see whether there is a critical level of shape–colour co-occurrence required to generate reliable extension behaviour.

[6] Piloting suggested that a lengthy Shapes block was required for participants to reliably exploit the shape–colour association in the Fixed Associations condition.

[7] Models were run in R (R Core Team, 2023) using lmer (Bates et al., 2015); successive difference coding used the contr.sdif function from the MASS package (Venables et al., 2002); plots were produced in ggplot2 (Wickham, 2009). For all models reported here the random effects structures consist of by-pair (not by-participant) random intercepts and random slopes for fixed effects that varied within-pair (e.g. block). Including a by-participant random effect nested within pair produced substantial convergence issues and explained very small amounts of variance. Data and analysis code is available at https://osf.io/btyus/?view_only=9fd9a85dafcd453baee315186374caac.

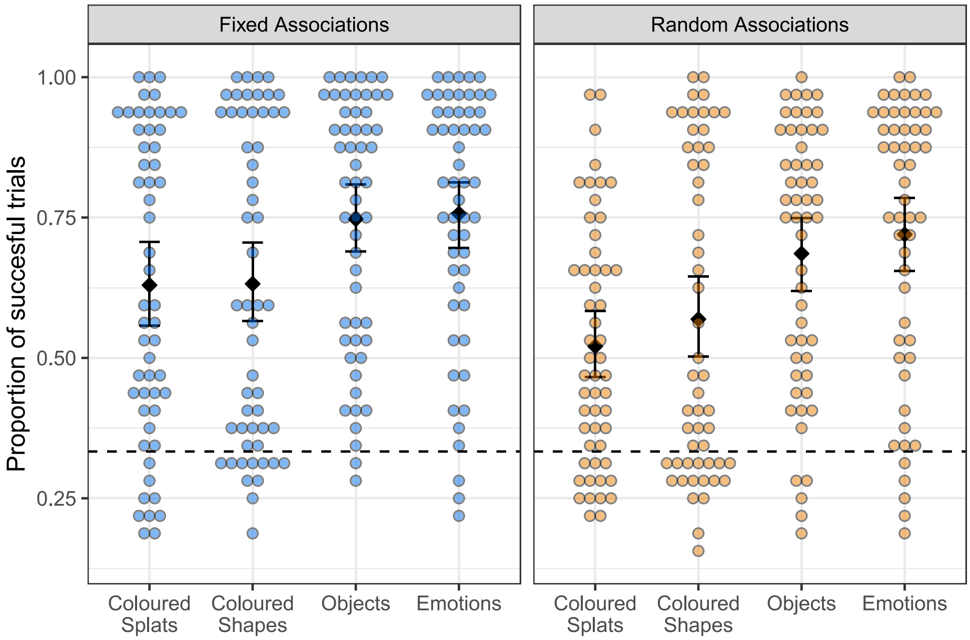

Fig. 2. Proportion of successful trials in the communicative task against block. Coloured points give means for individual pairs, black diamonds plus error bars indicate means of pair means plus bootstrapped 95% CIs. Dashed lines gives chance performance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

suggest that the presence of reliable statistical associations between colour and shape established in the Shapes block in the Fixed Associations condition facilitate the initial metonymic extension of those labels relative to the Random Associations baseline, and that advantage of early shared associations persists in the subsequent chain of extensions.

2.2.2. *Patterns of extension of meaning across blocks*

In the Fixed Associations condition, we expected that participants would extend the meaning of a given label to include the colour associated with that shape in the Shapes block (e.g. if stars were always red in the Shapes block, they should extend the meaning of the star label to encompass the meaning 'red' in the Coloured Splats block). Participants in the Random Associations condition could have also done the same — due to the random colour–shape associations there will tend to be some shapes which occur more often with a particular colour — but since those associations are weaker we expected the extension behaviour in this condition to be more arbitrary (as also suggested by the lower performance in the communicative task in the extension blocks shown in Fig. 2). Beyond the initial extension in the Coloured Splats block, we expected that participants in both conditions would extend their labels for colour in predictable ways, based on the associations they had established: e.g. having established a label for 'red' in the Coloured Splats block, they should continue to use this label in the Coloured Shapes block; they should then extend this label metonymically in the Objects block to refer to 'volcano', and metaphorically in the Emotions block to refer to 'anger'.

We evaluate this by coding, for each trial in each extension block, whether the label used reflected an extension as predicted (coded as 1) or an extension not as predicted (coded 0), yielding data which can be analysed using logistic regression.[8] The predicted label depends on the experimental condition and on the particular semantic extension being investigated, but is always the label most frequently associated with a colour in an earlier block.[9] We assume that the shape–colour associations made in the Shapes block are the basis for extensions in the Coloured Splats block; that the shape–colour associations made in the Coloured Splats block are the basis for extensions in the Coloured Shapes block; and that the shape–colour associations made in the Coloured Shapes block are the basis for extensions in the Objects and Emotions blocks. For example, if most of the pink shapes in the Shapes block are hexagons, then every use of the hexagon label to refer to the pink splat in the Coloured Splat block is coded as a predicted extension, and every use of a different label to refer to the pink splat is coded as a non-predicted extension.

Fig. 3 plots this block-to-block extension measure. The proportion of predicted extensions in the Shapes-Coloured Splats transition is reliably higher in the Fixed Associations than Random Associations condition, as expected (b=0.86, SE=0.17, p<.001): participants in the Fixed Associations condition generally exploit the associations between colour and shape provided in the Shapes block; in contrast, the proportion of predicted extensions in the Random Association condition cluster around 0.25, reflecting colour–label associations in the Coloured Splats block which are effectively arbitrary with respect to any colour–shape associations present in the Shapes block (random extension, assuming a unique shape for each target concept, would lead to around 25% of extensions being as predicted).

In subsequent blocks the two conditions look more similar: proportion of predicted extensions for the Coloured Splats-Coloured Shapes transition does not differ significantly between conditions (b=0.31, SE=0.18, p=.089, obtained from a model excluding the Shapes-Coloured Splats transition where block transition is treatment coded). Those associations are often extended in the predicted way to the Objects block, although we see a progressive weakening of the correspondence

[8] We also ran an additional analysis using the KL-divergence between the signal-meaning co-occurrence frequencies at successive blocks as a measure of whether labels were extended as predicted. This analysis and its results are detailed in the analysis code available at https://osf.io/btyus/?view_only=9fd9a85dafcd453baee315186374caac, and produces the same pattern of significant effects; we prefer the logistic regression analysis on binary data here since it is simpler.

[9] In the event of ties in association frequencies in the relevant block, the predicted label is chosen at random from the equally-frequent options for that colour (e.g. if pink co-occurs with hexagons and squares with equal frequency, either hexagon or square is selected at random as the predicted label for 'pink').

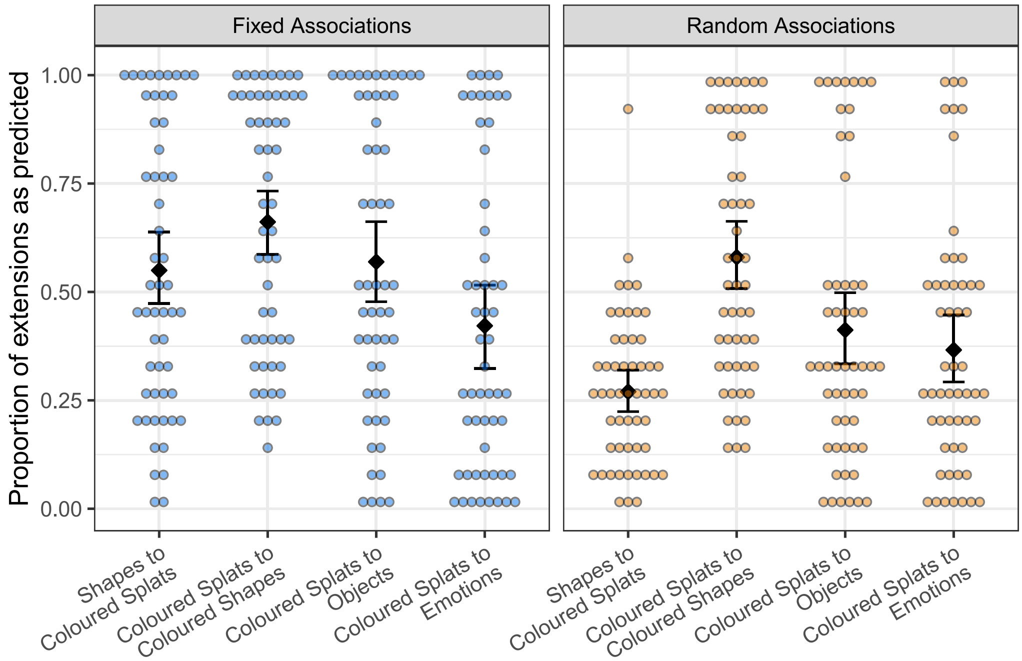

Fig. 3. Extension behaviour, as proportion of predicted extensions. Plotting conventions are as in Fig. 2.

in the extension to Objects and particularly in the extension to Emotions (proportion of predicted extensions decreases at successive comparisons: Coloured Shapes-Objects predicted extensions are lower than Coloured Splats-Coloured Shapes, b=−0.72, SE=0.20, p<.001; Coloured Shapes-Emotions predicted extensions are lower than Coloured Shapes-Objects, b=−0.80, SE=0.26, p=.002, obtained from a model with block transition coded using successive differences). In the Emotions block many pairs have proportions of predicted extensions at or around 0.25, indicating labelling conventions for emotion which do not extend labels in the predicted direction, although 13 pairs in the Fixed Associations and 7 in the Random Associations condition do have high (greater than 0.75) proportions of predicted extensions, indicating the conventions are extended in the predicted way in some pairs.[10] The gradual departure from predicted extensions does not differ significantly between conditions, as indicated by non-significant condition x block interactions in this model (interaction between block and Coloured Shapes to Objects: $b = 0.39$, SE=0.20, p=.052; interaction between block and Coloured Shapes to Emotions: b=−0.43, SE=0.25, p=.083; a significant effect with a positive sign would indicate less reduction in predicted extensions in the Fixed Associations condition, a negative sign would indicate greater reduction in the Fixed Associations condition).

In summary, this analysis suggests that the Fixed Associations condition shows the predicted pattern of extension of the statistical associations established in the Shapes block to the Coloured Splats block, whereas pairs in the Random Associations condition have to establish colour–shape correspondences from scratch. We find no difference between conditions in the subsequent blocks: pairs in both conditions extend those associations, with declining fidelity, to the subsequent blocks, and many participants in the Emotions block establish new emotion-shape associations which are unrelated to their established colour–shape associations. However, despite this tendency to drift away from the expected pattern of extensions in later blocks, in all blocks extending label meanings in the predicted way does facilitate communication. In each block for each pair we calculated that pair's proportion of predicted extensions and proportion of successful communication trials, and correlated these two numbers, producing positive correlations throughout: in the Coloured Splats block, $r = 0.68$, p<.001 (for the Coloured Splats block we only looked at pairs in the Fixed Associations condition, since the basis for the predicted extension is by design weak in the Random Associations condition); in the Coloured Shapes block, $r = 0.94$, p<.001 (not extending the established colour–shape associations in this block would be particularly challenging for communication); in the Objects block, $r = 0.49$, p<.001; in the Emotions block, $r = 0.22$, p=.021.

2.2.3. Non-arbitrary colour–shape correspondences

We were surprised that participants in the Random Associations condition were so successful at bootstrapping a signalling system for colour, leading to a smaller-than-expected (but still clear and significant) difference between conditions in the Coloured Splats block. Table 1 shows the frequency with which Senders in the Random Associations condition used each label for each colour in the Coloured Splats block. There are clear preferences for certain colours to be conveyed using particular shapes, which we had not anticipated when designing our stimuli and which pairs were able to exploit to rapidly coordinate on semantic extensions. Some of these are in hindsight obvious (yellow is strongly associated with 'star'), some have a conceivable retrospective justification (red seems to be associated with 'cross'; this could be due to the emblem of the Red Cross, the international humanitarian charity, or the fact that red and crosses are both associated with prohibition), and some are more baffling (grey is associated with circles and squares, although only a philistine would consider those to be boring shapes). Only pink seems to lack a clearly preferred shape. These impressions receive support from a simple χ^2 test of independence run on the tables of associations or individual rows from that table: the overall table shown in Table 1 is significantly non-uniform (p<.001), as are all of the rows associated with each colour apart from pink (p=.057 for pink; p<.001 elsewhere).[11]

[10] See also the KL-divergence analysis at https://osf.io/btyus/?view_only=9fd9a85dafcd453baee315186374caac, which provides a z-score for each participant which can be interpreted as reflecting a per-participant probability that their extension behaviour is by chance rather than as predicted.

[11] Note that this data violates the χ^2 assumption of independence, since each pair of participants contributes multiple data points; however, we consider this sufficient evidence that such associations probably exist.

Table 1
Counts of colour–label correspondences in the Coloured Splats block across all pairs in the Random Associations condition in Experiment 1 (highest association in each row in bold). Recall that every pair was allocated only 4 of these 7 shape labels.

	○	✚	⬡	☆	◇	⬠	□
Grey	**97**	42	69	38	36	59	83
Pink	60	68	62	39	**75**	62	58
Red	69	**106**	40	51	34	40	84
Yellow	67	69	30	**132**	36	53	37

Furthermore, using these non-arbitrary associations seems to improve communicative success in the Coloured Splats block in the Random Associations condition. We coded each trial for whether the sender had used a non-arbitrary label for the target, i.e. one of the most salient non-arbitrary associations in Table 1 (grey-circle, pink-diamond, red-cross, yellow-star), and ran a logistic regression predicting communicative success based on non-arbitrariness of the colour–label mapping used by the Sender (non-arbitrariness sum-coded, by-pair random intercepts, models with by-pair random slopes for non-arbitrariness did not converge but showed the same effects). The effect of using a non-arbitrary label was positive and significant (b=0.42, SE=0.07, p<.001).

One possible additional consequence of the presence of these non-arbitrary associations, pointed out to us by a reviewer, is that they may interfere with the establishment of arbitrary conventions in the Fixed Associations condition: the arbitrary associations between colour and shape established in the Shapes block in that condition might compete with these non-arbitrary associations for senders or receivers, impeding convergence on either mapping. We therefore repeated the analysis of the effect of (non-) arbitrariness on success on the Coloured Splats block in the Fixed Associations condition, reversing the coding such that each sender production was coded for arbitrariness rather than non-arbitrariness; the effect of arbitrariness on communicative success was negative (as predicted by the reviewer) but not significant (b=−0.10, SE=0.12, p=.369[12]), suggesting little support for this hypothesis in our data. In other words, the presence of these non-arbitrary associations seems to facilitate communication in the Random Associations condition in the absence of any other grounding for a colour–shape convention, but does not block the successful exploitation of statistically-grounded associations in the Fixed Associations condition.

The same reviewer also suggested that participants in the Fixed Associations condition who do not extend as predicted in the Coloured Splats block, i.e. fail to exploit the associations between shape and colour established in the Shapes block, might also show evidence of exploiting these same non-arbitrary associations. There is support for this hypothesis in our data. We analysed the Sender data from the Coloured Splats block in the Fixed Associations condition, coding each label selected by the sender as non-arbitrary (i.e. grey-circle, pink-diamond, red-cross, yellow-star) or not, and ran a logistic regression to test whether use of a non-arbitrary label was conditioned on whether their extension was as predicted or not (the fixed effect of as-predicted was dummy coded such that the model intercept reflects probability of a non-arbitrary label choice in cases where the participant extended as predicted, the model included by-pair random intercepts and random slopes for predictedness). As expected this model indicated a strong tendency not to use non-arbitrary mappings for predicted extensions (b=−4.0, SE=0.78, p<.001), but also a significantly increased production of non-arbitrary labels when extension was not as predicted by the associations in the Shapes block (b=2.19, SE=0.88, p=.013; note however that these non-arbitrary productions are still in the minority), confirming the reviewer's expectation and the unanticipated salience of these non-arbitrary associations.

2.2.4. Non-arbitrary emotion-shape correspondences

Similarly, the rather surprising level of communicative success achieved in the Emotions block in both conditions, despite the quite strong tendency not to extend existing labels in the predicted way, might reflect the presence of competing non-arbitrary emotion-shape correspondences. Table 2 shows the frequency with which Senders used each label for each emotion in the Emotions block, in both conditions but restricted to trials where the predicted extension was not used (e.g. the label used for red in the Coloured Shapes block was not extended to anger). As for the analysis of colour–shape correspondences, there are clear preferences for certain emotions to be conveyed using particular shapes, and these impressions receive support from a simple χ^2 test of independence run on the tables of associations or individual rows from that table: the overall table shown in Table 1 is significantly non-uniform (p<.001), as are all of the rows associated with each emotion (p<.001), although note that the emotions-shape association table is overall less skewed than the colour–shape association table (as indicated by a comparison of χ^2 values after conversion to proportions to account for the different amounts of data, the emotion-shape co-occurrence matrix is more uniform than the colour–shape matrix: Table 1: $\chi^2 = 208.4$; Table 2: $\chi^2 = 172.1$, where higher values indicate greater departure from uniformity).

For some of these associations we can offer a plausible explanation leveraging pre-existing associations. Shapes with lots of corners and sharp angles are associated with anger or excitement, whereas rounded shapes with curving edges are associated with contentment or sadness (Sievers et al., 2019). The former is consistent with the associations of love and star, anger and cross in our data; we would note that stars are generally positive, as true love should be, and that crosses are associated with prohibition and prohibition is associated with anger (and also that "cross" is a synonym of "angry" in UK English, although we are skeptical that many participants were exploiting this potential pun). The association between roundedness and contentment noted by Sievers et al. (2019) is consistent with the association in our data between happiness and circles. Other associations are far more mysterious, and we can offer no plausible explanation why sadness might be strongly associated with hexagons, or happiness with squares as a close second to circles.

However, as for the non-arbitrary colour–shape associations discussed in the previous section, there was evidence of an effect of the use of these non-arbitrary emotion-shape associations on communicative success. As in the equivalent analysis of colour–shape associations, we coded each trial in the Emotions block for whether the sender had used a non-arbitrary signal for the target, i.e. one of the most salient non-arbitrary associations in Table 2 (sad-hexagon, love-star, angry-cross, happy-circle; note that we consider data from both the Fixed and Random Associations conditions in this analysis, since by the time they reach the Emotions block both sets of participants have established colour–shape associations that could be extended to emotions), and ran a logistic regression predicting communicative success based on non-arbitrariness of the emotion-label mapping used by the Sender (non-arbitrariness sum-coded, by-pair random intercepts and random

[12] This model converged with by-pair random slopes for arbitrariness but the same model lacking these random slopes showed the same n.s. effect, i.e. the absence of a significant effect in the Fixed Associations condition does not depend on the use of a more conservative model.

Table 2
Counts of emotion-label correspondences in the Emotions block across all pairs in Experiment 1, for non-predicted extensions only (highest association in each row in bold). Recall that every pair was allocated only 4 of these 7 shape labels.

	○	✚	⬡	☆	◇	⬠	□
Sad	55	75	**104**	30	65	78	79
Love	96	106	38	**136**	49	52	83
Angry	66	**105**	77	63	39	97	58
Happy	**112**	50	43	82	68	57	111

slopes for non-arbitrariness). The effect of using a non-arbitrary label was smaller than that seen for colour–shape associations but still positive and significant (b=0.18, SE=0.09, p=.042). Similarly, we also find that participants who do not extend as predicted in the Emotions block also show evidence of exploiting these same non-arbitrary associations. As for the equivalent colour–shape analysis, we analysed the Sender data from the Emotions block, coding each label selected by the sender as non-arbitrary according to the associations seen in Table 2 or not, and ran a logistic regression to test whether use of a non-arbitrary label was conditioned on whether their extension was as predicted or not (same model structure as the equivalent analysis for colour–shape). As expected this model indicated a strong tendency not to use non-arbitrary mappings for the predicted extensions (b=−3.52, SE=0.58, p<.001), but also a significantly increased production of non-arbitrary labels when extension was not as predicted (b=1.44, SE=0.65, p=.026; note however that these non-arbitrary productions are still in the minority).

Our interpretation of these findings is that a combination of extending some existing labels in the expected way, leveraging other (sometimes mysterious) non-arbitrary associations between emotions and shape, and rapidly adapting to a partner's choices, allows successful communication in the Emotions block; note, however, as shown above, that making extensions as predicted (i.e. using the established label for red to convey anger, the established label for yellow to convey happiness, and so on) is positively correlated with communicative success in the Emotions block (r=0.22, p=.021).

2.3. Experiment 1 discussion

Our results indicate that participants were able to draw on both metonymic and metaphorical associations in extending the meaning of labels to apply to colours, and then subsequently to more concrete meanings (objects) and more abstract meanings (emotions). Extension via grounding in metonymic and metaphorical associations has been argued to form an important mechanism in both the early evolution of linguistic systems (e.g. Smith & Hoefler, 2015), and in subsequent language change (e.g. Traugott & Dasher, 2001). Participants used a range of different sources of information to establish the meaning of the symbols. Our results suggest that semantic extension is facilitated by the presence of metonymic associations (in our Fixed Associations condition) that are part of the shared perceptual environment for a pair, presumably since these are reliably part of their common ground and therefore highly likely to be grasped by the audience, and this advantage persists across subsequent extensions of those labels. However, even without such salient associations, our participants were able to exploit other metonymic associations that are plausibly part of common ground (e.g. regarding the colours of stars and crosses) in order to converge on a working communication system.

One additional finding, clearest in the Emotions block, is that the degree to which a grounding can be assumed to be common ground has an important influence on which grounds will be exploited in communication. This can be seen in the contrast between the Objects and Emotions blocks. The colour of an object is broadly uncontroversial: we can be confident that even an entirely unknown interlocutor will know that e.g. bananas are yellow. However, associations between colours and emotions are likely to be more subjective and therefore less likely to be shared between interlocutors (e.g. while love is typically associated with pink, associating love with tumultuous passion could lead to an individual associating love with red), which could explain the decreased tendency to rely on exploiting established colour–shape associations in the predicted way in the Emotions block as compared to the Objects block; if in a given pair each person had different colour-emotion associations, they would have had to negotiate emotion-shape correspondences from scratch. Other emotion-shape correspondences may also have provided an alternative means of grounding labels for emotions; we see evidence of several such associations in our data, including between anger and the cross shape, or love and the star, and using these associations seems to be positively associated with communicative success in cases where participants chose not to extend established colour–shape associations in the way we predicted. This also suggests that the specific nature of the grounds for semantic extension may be less important than the fact that there is a motivation for semantic extension, which is sufficiently easily accessible to the interpreter (or deemed so by their interlocutor) for them to be able to work out the intended extension.

However, there is at least one potential difference between the Objects and Emotions block in Experiment 1 which might explain the reduced tendency to rely on established colour–shape associations as the basis for metaphors in the Emotions block: whereas the Objects block immediately follows the Coloured Shapes block, meaning those colour–shape associations are presumably highly salient for participants, the Objects block intervenes between the Coloured Shapes and Emotions block. This means that colour–shape associations may be less salient at the Emotions block, and/or additional unhelpful associations established in the Objects block may interfere with the predicted extension of colour terms to emotions (e.g. participants who have established an association between say the hexagon symbol and the colour pink may be reluctant to extend the hexagon to the emotion 'love' given that it has more recently been used to mean 'pig'). In Experiment 2 we address this by running a version of the experiment where the extensions which are made within-subjects across multiple sequential extensions in Experiment 1 are instead between-subjects in Experiment 2: after the Coloured Splats block participants completed a single extension block, extending either to Coloured Shapes, Objects, or Emotions. Additionally we added a scrambling manipulation on the shapes seen by each participant, in an attempt to reduce the utility of the unanticipated shape–colour and shape–emotion associations we saw in Experiment 1 (i.e. star-yellow, cross-red) and therefore sharpen the differences between the Fixed and Random Associations conditions.

3. Experiment 2

As in Experiment 1, each pair was randomly allocated either to the Fixed or Random Associations condition, determining whether they encountered deterministic or variable associations between shapes and colours in the Shapes block of the experiment. Each pair was additionally assigned a random extension block to complete: either Coloured Shapes, Objects, or Emotions. This yields a 2 *x* 3 between-pairs design where we manipulate reliability of associations in the Shapes block (2 levels) and the type of extension block (3 levels) each pair completed.

3.1. Methods

3.1.1. Participants

820 English-speaking adults were recruited through the online crowdsourcing platform Prolific. We obtained complete data from 352 pairs (704 participants), 169 pairs in the Fixed Associations condition (56 in the Coloured Shapes extension condition, 54 in the Objects extension condition, 59 in the Emotions extension condition) and 183 pairs in the Random Associations condition (58 Coloured Shapes, 61 Objects, 64 Emotions). As in Experiment 1, we used Prolific's filters to restrict the experiment to people who self-identified English as their first language, and we did not use any geographical restrictions; 70% gave the UK as country of birth (8% gave South Africa, 4% gave Ireland and Canada, 2% gave Australia, USA and Zimbabwe, all other countries of birth were represented at 1% or less each), and 74% gave the UK as their country of residence (10% South Africa, 4% Canada, 3% Ireland, 2% Australia and USA, other countries of residence were represented at 1% or less each). Participants who completed the experiment (median completion time 27 min) were paid £5; partial payment was made to participants who were unable to complete the experiment (e.g. due to not being paired with a partner).[13]

3.1.2. Materials

The materials were the same as in Experiment 1.

3.1.3. Procedure

The infrastructure for the experiment was the same as in Experiment 1 (JavaScript plus jsPsych at the participant side, WebSockets and a python server coordinating pairs of participants), with minor edits to handle the change from a within- to between-pairs treatment of extension blocks. Full code for the experiment is available at https://github.com/kennysmithed/SemanticExtension.

The procedure for Experiment 2 was largely the same as for Experiment 1: after the pre-screening for colour-blindness, warm-up communicative task, and pairing with a partner, participants worked through a trivial White Shapes block (8 trials), followed by the Shapes block (64 trials) which allowed us to established Fixed or Random associations between colours and shapes, then a Coloured Splats block (32 trials) where participants were required to use shape to communicate about colour. However, after completing the Coloured Splats block, pairs then progressed to their 4th and final block, a second extension block (32 trials) which involved extension to either Coloured Shapes, Objects, or Emotions (randomly allocated per pair) — this second extension block was identical to the equivalent block in Experiment 1, but for the Objects and Emotion blocks immediately followed the Coloured Splats block, unlike in Experiment 1.

As an additional change for Experiment 2, we remapped the shapes that each participant in a pair saw: each *participant* (rather than pair) was allocated 4 shapes selected at random from the set of 7 options (square, circle, diamond, star, cross, pentagon, hexagon), and then within a given pair those shapes were mapped randomly but consistently to one another, such that whenever participant 1 in the pair saw a particular shape, shape A, their partner participant 2 would see their corresponding shape A. The remapping process is illustrated in Fig. 4. Remapping was applied exhaustively, including to the receiver array, to the shape label received from the partner, and to the shapes and labels seen during feedback, ensuring that the manipulation was invisible to individual participants. The intention was to reduce participants' tendency to rely on pre-existing shape–colour (or shape–object or shape–emotion) correspondences, since what was intuitive for one participant (e.g. star-yellow) would be arbitrary for their partner (star remapped to square yields a square-yellow association for the partner).

[13] 27 participants admitted to taking written notes as part of the post-experiment debrief and were excluded from the analysis, together with their partners, but paid in full.

3.2. Results

3.2.1. Communicative success

As in Experiment 1, communicative success (the proportion of trials where the Receiver clicked on the correct response) in the first two blocks of the experiment (White Shapes, Shapes) was uniformly high, with average accuracy of over 99% in both conditions. Fig. 5 shows communicative success on the subsequent blocks of the experiment. Participants were quite successful in the task in both conditions, and many pairs were able to reach high levels of accuracy in using shapes to communicate about colour (in the Coloured Splats blocks) and could successfully extend those meanings again as required in their second extension block. Again, participants in the Random Associations condition do however appear to be at a disadvantage in this extension process due to the lack of shared statistical associations between colour and shape in the early stages of the experiment, particularly in the Coloured Splats block, the first block where they have to creatively extend labels to convey colour rather than shape.

As in Experiment 1, we used logistic mixed effects models to analyse the binary outcome of communicative success on each trial (success/failure), with condition, block, extension type (in the final extension block: coloured shapes, objects, or emotions) and their interaction as fixed effects. A model with block treatment-coded (with the Coloured Splats block as the reference level) and condition (Fixed or Random Associations) sum-coded shows that the Fixed and Random Associations conditions differed at the Coloured Splats block ($b=0.24$, $SE=0.07$, $p<.001$), with participants in the Fixed Associations condition communicating more successfully, as in Experiment 1. Performance on the second extension block (block 4 of the experiment) was higher ($b=0.49$, $SE=0.06$, $p<.001$) on average, with the increase in the Fixed Associations condition being smaller ($b=-0.12$, $SE=0.06$, $p=.042$) — this result suggests that, as in Experiment 1, the presence of reliable associations between colour and shape established in the Shapes block in the Fixed Associations condition gives pairs in that condition an advantage in initially extending the meaning of those labels, and that advantage largely persists in a second extension. Although there are some fluctuations in the success levels at the second extension block depending on whether the second extension was to Coloured Shapes, Objects or Emotions, none of the interactions between block and extension type were significant (lowest p value seen in the interaction between block and extension to Coloured Shapes: $b=-0.16$, $SE=0.08$, $p=.066$), indicating that once performance in the Coloured Splats block was controlled for, participants were largely equally successful at extending colour–shape associations established in the Coloured Splats block to Coloured Shapes, Objects or Emotions.

3.2.2. Patterns of extension of meaning across blocks

Fig. 6 plots the proportion of extensions made using predicted labels, which are defined as for Experiment 1, except that the shape–colour associations made in the Coloured Splats block always form the basis for the predicted labels in the second extension, due to the revised between-pair design of the extension blocks in Experiment 2. The proportion of predicted extensions in the Shapes-Coloured Splats transition is reliably higher in the Fixed Associations than Random Associations condition, as expected and as in Experiment 1 ($b=1.03$, $SE=0.13$, $p<.001$): participants in the Fixed Associations condition generally exploit the associations between colour and shape provided in the Shapes block; in contrast, the proportion of predicted extensions in the Random Association condition cluster around 0.25, reflecting colour–label associations in the Coloured Splats block which are effectively arbitrary with respect to any colour–shape associations present in the Shapes block.

In the subsequent extension to either Coloured Shapes, Objects or Emotions the Fixed and Random Associations conditions look more similar, but there is a clear effect of the nature of the second extension, with fewer pairs extending as predicted in Objects and particularly

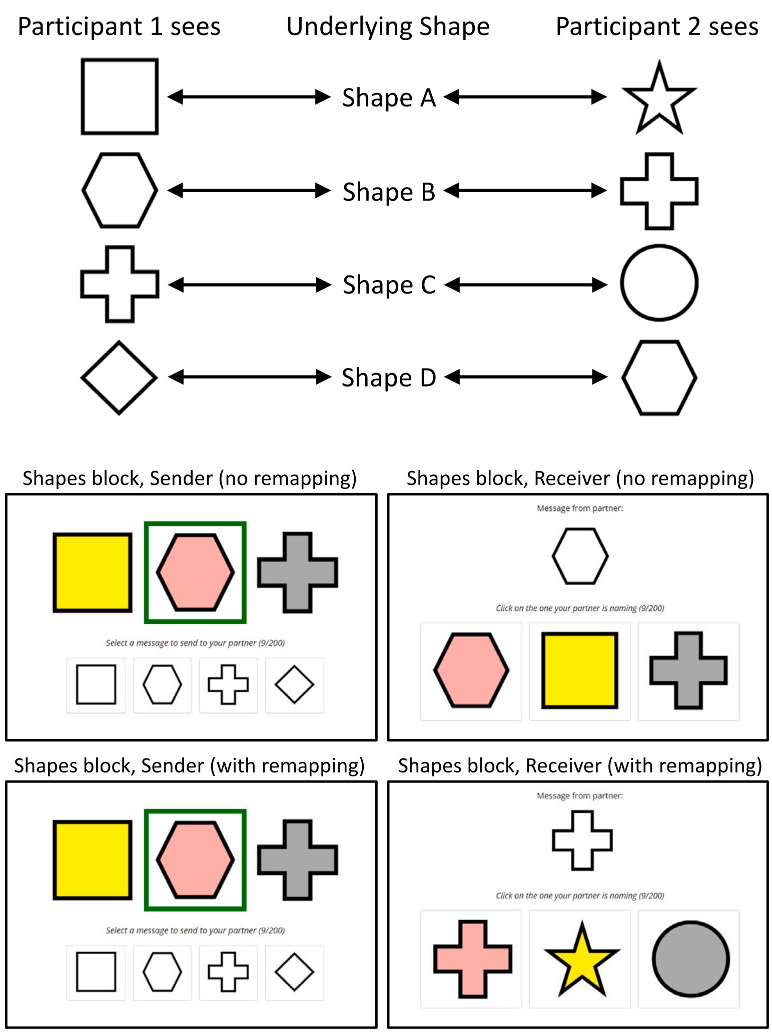

Fig. 4. An example of the remapping process applied in Experiment 2. Upper: Each participant receives their own random allocation of 4 shapes, which are remapped consistently to their partner's shapes. Middle: in Experiment 1 (where no remapping was applied) both participants saw the same shapes when acting as sender and receiver. Lower: in Experiment 2, the remapping process means that participants see a randomly but consistently remapped shape — here, we apply the remapping from the upper panel to the example in the middle panel and assume that Participant 1 is the sender. Participant 1 chose to send a hexagon symbol (Shape B) to communicate a pink hexagon target; Participant 2 is the receiver and receives a (remapped) cross symbol, allowing them to correctly select the pink cross from their array.

Emotions. An analysis focusing on predicted extensions in this second block and using successive difference coding for the type of second extension block shows that the proportion of predicted extensions for the Coloured Shapes extension block differs between Fixed and Random Associations conditions (b=0.25, SE=0.10, p=.013), with Fixed Associations pairs making more predicted extensions than pairs in the Random Associations condition, an effect that was not apparent in Experiment 1. The model also shows that participants make fewer predicted extensions in Objects than in Coloured Shapes (b=−0.84, SE=0.25, p<.001), and yet fewer predicted extensions in Emotions than in Objects (b=−0.79, SE=0.24, p=.001); the absence of interactions with Fixed vs Random Associations (smallestp = 0.7) indicates both that the higher proportion of predicted extensions for the Fixed Associations condition seen in the Coloured Shapes extension is also present in the extension to Objects and Emotions, and that this declining tendency to extend established colour–shape correspondences to Objects and Emotions does not differ based on the reliability of the shape–colour associations established early in the experiment. Crucially, Experiment 2 therefore matches the result from Experiment 1, that participants less consistently extend established colour–shape associations to objects and, particularly, emotions — in other words, this was not an artefact of the prolonged sequence of extensions participants were required to make in Experiment 1. In fact, strikingly few participants extend their shape–colour associations to emotions in the predicted way in Experiment 2, as evidenced by the scarcity of participants who have predicted extension proportions above 0.75 in the Emotions block.[14] This suggests that the prolonged use of colour–shape associations and their repeated use for extension to objects and then emotions in Experiment 1 might in fact have favoured extension as predicted, rather than inhibiting it. The pattern of correlations between extension-as-predicted and communicative success in Experiment 2 mirrors that seen in Experiment 1, despite the change in design: as for Experiment 1, in each block for each pair we calculated that pair's proportion of predicted extensions and proportion of successful communication trials, and correlated these two numbers, producing positive correlations everywhere, although this correlation is not significant in the Emotions block: in the Coloured Splats block, $r = 0.75$, p<.001 (looking at pairs in the Fixed Associations condition only); in the Coloured Shapes block, $r = 0.90$, p<.001; in the Objects block, $r = 0.37$, p<.001; in the Emotions block, $r = 0.13$, p=.14 (this was the weakest correlation, but still significant, in Experiment 1).

[14] Again, see the KL-divergence z-scores in the supporting online analyses for interpretable z-score values per participant.

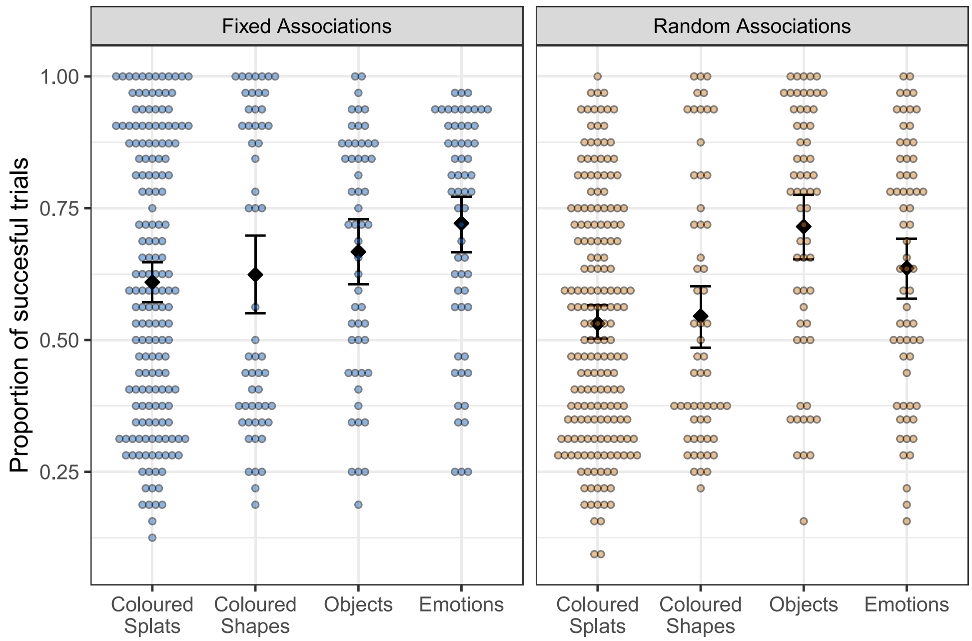

Fig. 5. Success in the communicative task against block in Experiment 2. We plot in the same style as Experiment 1 for ease of comparison, but note that in Experiment 2, all pairs completed the Coloured Splats block, and then moved on to *one* of Coloured Shapes, Objects, or Emotions, completing this block only. Plotting conventions are as in previous plots. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

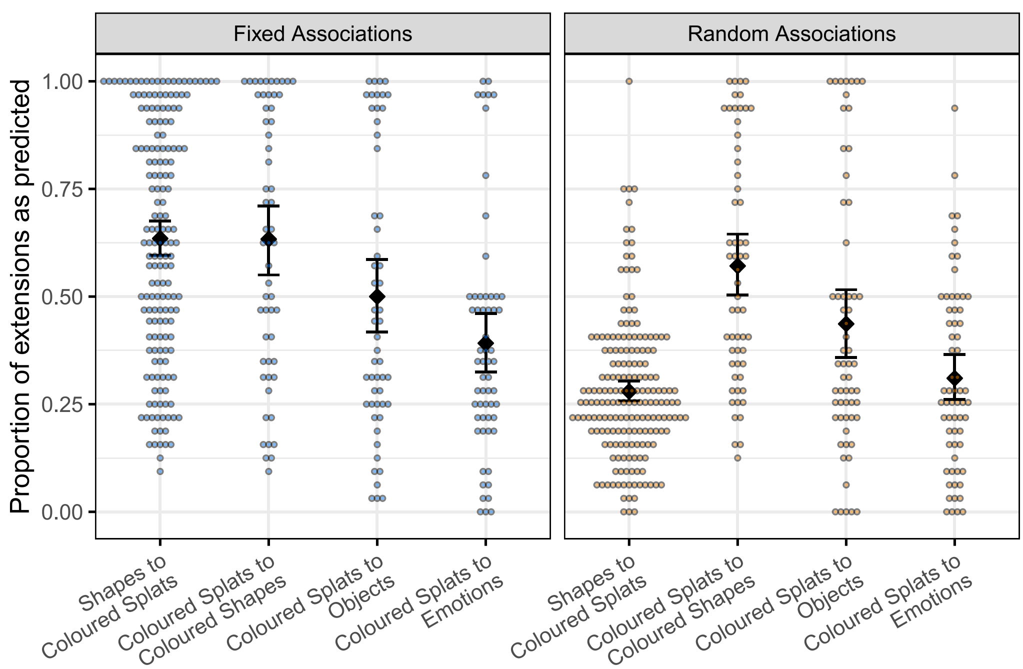

Fig. 6. Extension behaviour, as proportion of predicted extensions. Plotting conventions are as in previous figures.

3.2.3. Non-arbitrary colour–shape correspondences

In Experiment 2, we attempted to suppress the ability of participants to exploit pre-existing non-arbitrary colour–shape (or colour-object or colour-emotion) associations, by allocating each participant a random set of shapes and then consistently remapping between those two idiosyncratic inventories.

To evaluate this manipulation, we compared performance between Experiments 1 and 2 on our two main measures (communicative success and the proportion of extensions made using predicted labels) in the Coloured Splats block; this is equivalent across the two experiments and represents in both cases the first block in which participants extend shape labels to communicate about colour. Surprisingly, we found that remapping participants' colour–shape associations had little impact on either measure: logistic regressions with experiment and condition as predictors and by-pair random intercepts showed both the main effect of the experiment and the condition *x* experiment interaction as not significant for both communicative success (main effect of experiment: b=−0.02, SE=0.07, p=.811; experiment *x* condition interaction: b=−0.04, SE=0.07, p=.545) and proportion of extensions as predicted (main effect of experiment: $b = 0.13$, SE=0.08, p=.088; experiment *x* condition interaction: $b = 0.9$, SE=0.08, p=.267).

However, there was evidence that the remapping reduced (but did not eliminate) participants' reliance on pre-existing, non-arbitrary colour–shape associations. Table 3 shows raw frequencies of the colour–label pairings used in the Coloured Splats block of the Random Associations condition in Experiment 2, corresponding to the equivalent frequencies for Experiment 1 shown in Table 1. Some associations from Experiment 1 are still clearly apparent in Experiment 2 (e.g. yellow-star remains prominent, suggesting that if either participant has a star available in their repertoire of labels they are likely to map it to 'yellow', and their partner is likely to simply adopt the corresponding remapped association, e.g. yellow-pentagon), while others (e.g. pink-diamond, grey-circle, red-cross) are less prominent. A comparison of χ^2 values for each table, after conversion to proportions to account for the different amounts of data, reveals that the co-occurrence matrix for Experiment 2 is more uniform than the equivalent for Experiment 1 (Table 1: $\chi^2 = 208.4$; Table 3: $\chi^2 = 62.3$, where higher values indicate greater departure from uniformity).

However, the colour–shape associations do correlate across experiments (as assessed by a rank correlation on the raw values in Tables 1 and 3: τ =0.49, p<.001), and there was still evidence of an effect of the use of the non-arbitrary associations identified in Experiment 1 on communicative success in Experiment 2. As in the equivalent analysis in Experiment 1, we coded each trial in the Coloured Splats block of the Random Associations condition for whether the sender had used a non-arbitrary signal for the target, i.e. one of the most salient non-arbitrary associations in Table 1 from Experiment 1 (grey-circle, pink-diamond, red-cross, yellow-star), and ran a logistic regression predicting communicative success based on non-arbitrariness of the colour–label mapping used by the Sender (non-arbitrariness sum-coded, by-pair random intercepts and random slopes for non-arbitrariness). The effect of using a non-arbitrary label on communicative success was smaller than that seen in Experiment 1 but still positive and significant (b=0.25, SE=0.05, p<.001). As hypothesised by the reviewer who suggested this analysis, we also saw a small but significant *negative* effect on communicative success of using arbitrary associations in the Fixed Associations condition (b=−0.13, SE=0.05, p=.021); this effect was not significant in Experiment 1, but our sample size is substantially larger here. In other words, the presence of these non-arbitrary associations, although somewhat muted by our shuffling manipulation, seems to facilitate communication in the Random Associations condition in the absence of any other grounding for a colour–shape convention, and poses a modest impediment to the formation of statistically-grounded associations in the Fixed Associations condition, contributing to the rather modest difference in communicative success between the two conditions in the Coloured Splats block.

We also followed up on the reviewer suggestion that participants in the Fixed Associations condition who do not extend as predicted in the Coloured Splats block, i.e. fail to exploit the associations between shape and colour established in the Shapes block, might also show evidence of exploiting these same non-arbitrary associations; there was evidence for this in Experiment 1. Following the same procedure as the equivalent analysis in Experiment 1, we analysed the Sender data from the Coloured Splats block in the Fixed Associations condition, coding each label selected by the sender as non-arbitrary (i.e. grey-circle, pink-diamond, red-cross, yellow-star) or not (NB this coding was conducted with respect to the colours and shapes seen by the Sender; the colours and shapes seen by the Receiver will often have been different), and ran a logistic regression to test whether use of a non-arbitrary label was conditioned on whether their extension was as predicted or not (same model structure as the equivalent analysis for Experiment 1). As expected this model indicated a strong tendency not to use non-arbitrary mappings for the extensions which followed the statistical associations from the Shapes block (b=−2.0, SE=0.11, p<.001), but also a significantly increased production of non-arbitrary labels when extension was not as predicted by the associations in the Shapes block (b=0.52, SE=0.14, p<.001; note however that these non-arbitrary productions are still in the minority), confirming the reviewer's expectation and the salience of these non-arbitrary associations.

3.2.4. Non-arbitrary emotion-shape correspondences

Table 4 shows the frequency with which Senders used each label for each emotion in the Emotions block, in both conditions but restricted to trials where the predicted extension was not used (e.g. the label used for red in the Coloured Shapes block was not extended to anger). As for the analysis of colour–shape correspondences, there are clear preferences for certain emotions to be conveyed using particular shapes, and these impressions receive support from a simple χ^2 test of independence run on the tables of associations or individual rows from that table: the overall table shown in Table 1 is significantly non-uniform (p<.001), as are all of the rows associated with each emotion (p<.001).

As for colour-shape correspondences, the most prominent emotion-shape correspondences in Experiment 2 are different from those seen in Experiment 1, but the two sets of associations do correlate (as assessed by a rank correlation on the raw values in Tables 2 and 4: τ=0.38, p=.005), and there is still evidence of an effect of the use of the non-arbitrary emotion-shape associations identified in Experiment 1 on communicative success in Experiment 2. As in the equivalent analysis in Experiment 1, we coded each trial in the Emotions block for whether the sender had used a non-arbitrary signal for the target, i.e. one of the most salient non-arbitrary associations in Table 2 (sad-hexagon, love-star, angry-cross, happy-circle; note that this coding was based on the label the receiver saw), and ran a logistic regression predicting communicative success based on non-arbitrariness of the emotion-label mapping used by the Sender (non-arbitrariness sum-coded, by-pair random intercepts and random slopes for non-arbitrariness). The effect of using a non-arbitrary label was positive and significant, and roughly the same size as that seen in Experiment 1 (b=0.15, SE=0.07, p=.027). Similarly, we also find that participants who do not extend as predicted in the Emotions block, also show evidence of exploiting these same non-arbitrary associations. We analysed the Sender data from the Emotions block, coding each label selected by the sender as non-arbitrary according to the associations seen in Experiment 1 or not, and ran a logistic regression to test whether use of a non-arbitrary label was conditioned on whether their extension was as predicted or not (same model structure as the equivalent analysis for Experiment 1). As expected this model indicated a strong tendency not to use non-arbitrary mappings for the predicted extensions (b=−2.42, SE=0.22, p<.001), but also a significantly increased production of non-arbitrary labels when extension was not as predicted (b=0.52, SE=0.26, p=.041; note however that these non-arbitrary productions are still in the minority).

3.3. Experiment 2 discussion

The results in Experiment 2 confirm our earlier results that participants were able to extend their colour–shape associations to communicate successfully about Coloured Shapes, Objects and Emotions. They also indicate that differences between the three types were not an artefact of the fixed order in which they were explored in Experiment 1, but were rather influenced by the degree to which the grounds for the extensions can be considered part of common ground by the participants: the most predicted extensions are made to Coloured Shapes, which relies on the specific associations which have just been established in the experiments; the intermediate level is the metonymic extension to Objects, which relies on shared understanding of widely-known stereotypes regarding the salient colour of objects; while the fewest predicted extensions are made via metaphoric extension to Emotions, which relies on more subjective and perhaps less reliably shared connections to specific colours.

4. Discussion

We show across two experiments that the presence of reliable statistical associations between colour and shape established in the Shapes

Table 3
Counts of colour–label correspondences in the Coloured Splats block across all pairs in the Random Associations condition (highest association in each row in bold, highest associations from the equivalent table for Experiment 1 in italics) in Experiment 2. Recall that every individual was allocated only 4 of these 7 labels.

	○	✚	⬡	☆	◇	⬠	□
Grey	*219*	224	216	179	153	224	**249**
Pink	231	175	218	**254**	*225*	203	158
Red	215	*260*	192	215	151	161	**270**
Yellow	198	190	179	***399***	160	146	191

Table 4
Counts of emotion-label correspondences in the Emotions block across all pairs in Experiment 2, for non-predicted extensions only (highest association in each row in bold, highest associations from the equivalent table for Experiment 1 in italics). Recall that every pair was allocated only 4 of these 7 shape labels.

	○	✚	⬡	☆	◇	⬠	□
Sad	89	97	*100*	55	92	110	**122**
Love	**134**	98	100	*101*	69	71	110
Angry	60	*96*	**110**	85	53	88	107
Happy	*98*	83	75	**117**	81	92	67

block in the Fixed Associations condition facilitate the communicatively-successful initial metonymic extension of those labels, and that advantage of those early shared associations persists in a subsequent chain of extensions, where the new colour meanings are further extended to convey objects and emotions. Both experiments show that the extensions of colour terms to objects and emotions leads to less use of existing established conventional meanings: for example, most participants in the Emotions block establish new emotion–shape associations which are unrelated to their established colour–shape associations. The challenge of these extensions is presumably due to the less certain grounding of those extensions in shared experience, or the availability of other competing salient shared associations.

One of our primary goals for this work is to present what we believe is a quite flexible and general experimental paradigm for studying semantic extension, inspired by similar work applied to testing the mechanisms responsible for typological and diachronic universals experimentally. Below we provide several suggestions for how this approach could be applied to follow up on questions derived from our findings here or the broader literature. But our findings from the current study also speak to assumptions from the existing literature on semantic change, and make several predictions about semantic change in natural languages that we would like to highlight.

Firstly, and reassuringly, we find (as widely assumed in the literature) that semantic extension is facilitated by and reflects shared associations, with extensions made on the basis of shared associations being more probable and more communicatively successful. While this is as expected, it is not always the case that widely-held assumptions derived from historical or typological studies can be supported experimentally — for instance, the assumption that asymmetries in grammaticalisation (where change proceeds almost exclusively from lexical to functional, concrete to abstract meaning, rather than the reverse) reflects underlying asymmetries in associations between those semantic domains has repeatedly proven difficult to reproduce in experimental studies testing for those associational asymmetries in controlled circumstances (Jäger & Rosenbach, 2008; Kapron-King et al., 2025), calling into question either the experimental methods used or the original assumption itself.

Secondly, several aspects of our results suggest additional hypotheses that could be investigated in historical or typological datasets. First and most obviously, we find that metonymic extension is more frequent than metaphorical extension — this is particularly clear in Experiment 2, where participants are more likely to extend existing labels with a conventionalised colour meaning to convey metonymically-associated objects than metaphorically-associated emotions. Although metaphoric extension has typically received more attention in the semantic change literature, there has been a recent trend to reassess the role of metonymy as a mechanism in its own right and also as a bridging mechanism involved in metaphorical change (e.g. Barcelona, 2003; Goossens, 1990; Hamilton, 2016; Kövecses, 2013), and our results suggest that there might also be interactional grounds to support the possibility that metonymic extensions are particularly probable or particularly accessible. Indeed, Brochhagen et al.'s (2023) finding that extensions observed in corpora of semantic change and colexification seem to be primarily based on associativity between concepts, rather than taxonomic or visual similarity, is consistent with this idea that metonymic (i.e. association-driven) rather than metaphorical (i.e. similarity-driven) extension is favoured. Our method could be adapted to provide a more direct test of this hypothesis. Arbitrary associations between concepts can be established in our method by e.g. co-presentation during the experiment, as we do here for colour and shape; using novel object referents or referents drawn from a known taxonomy (e.g. as in Xu & Tenenbaum, 2007) would allow us to pit these various grounds for extension against one another, to explore whether there is an inherent preference for certain grounds (e.g. for association-based extensions over extensions based on visual similarity of referents).

Finally in our discussion of the implications of our results for understanding semantic extension in natural languages: there is some evidence in our data that chaining of semantic extensions is challenging — in general we see a reduction in predicted extension on the second or subsequent extension our participants make. This stands in contrast to Ramiro et al. (2018), who fit several models of semantic change to historical corpora and find that the best-fitting model of semantic change is one which assumes that all existing senses of a word are equally acceptable targets for extension. We suspect that this difference is due to the difference in timescales — our participants are attempting to chain on top of recently-established and rather ad-hoc extensions, whereas due to the nature of their data Ramiro et al. are studying chaining on top of extensions which are sufficiently well-established and conventionalised to appear in a historical thesaurus. Nonetheless, it could be that their model would be improved by assuming that the most recently-established senses are somewhat less likely to be targets for immediate extension; alternatively, our work highlights the importance of understanding the process and timescale of conventionalisation of one extension before it can form the basis for a subsequent extension.

Turning to obvious avenues for future work: our results complement those of Verhoef et al. (2016), who find that their participants reliably exploit a shared association between temporal duration and spatial extent to convey differences of temporal duration. Whereas this presumably derives from reliable associations between spatial extent and duration common to all participants, our method allows us to manipulate the associations that form the basis for these early conventions. Our method for studying chained semantic extensions could also be applied in the Verhoef et al. paradigm, for example to investigate semantic extension of temporal terms to markers of causality.

A reviewer points out that pre-existing shared associations between colours, emotions and shapes play an important role in our experiment, on top of the experimentally-manipulated shared associations between colour and shape we established in the Shapes block, and that these associations might differ between individuals and/or across cultures. We agree, but note that our participants were relatively homogeneous in this respect (see the participant demographics provided in the Participants sections of the Methods; over the two experiments, 69% of our participants give the UK as their country of birth, and 74% give the UK as their country of residence), meaning that the influence of cross-cultural variation in these associations is likely to be relatively small. Future work could use the same paradigm to test the same effects in individuals from different populations with (potentially) different associations, or deliberately construct dyads from groups with heterogeneous associations to manipulate the influence of pre-existing associations on grounding semantic extension.

An important feature of our paradigm is that participants are *forced* to extend existing labels to cover new concepts — their set of available labels remains fixed even as the set of concepts to be conveyed expands. An important extension to our method would be to allow participants to introduce new signals to convey new concepts, allowing them to trade off potential confusion or ambiguity arising from extension of an existing label versus the cost of establishing a convention featuring an entirely new label. Our expectation is that when the basis for the extension is transparent and likely to be shared, and context disambiguates between the word's original and extended meaning, participants will favour extension rather than using a new word, not least because it minimises the cognitive cost of the lexicon as a whole (e.g. Piantadosi et al., 2012). In contrast, where the grounding is less clear or the re-use of an existing label would generate in-context ambiguity, we expect that participants might strategically favour establishing new labels to convey new concepts. Indeed we already see results compatible with this in our existing experiment, where the 'obvious' extensions are made in the Coloured Splats block for the Fixed Associations condition, but many participants go on to establish entirely new conventions from scratch in the Emotions block, where the grounding of the extension of colour terms is less clear.

Returning to the issue of metonymic versus metaphorical extension, an intriguing observation from Brochhagen et al. (2023) is that child overextensions make more extensive use of grounds other than associativity (e.g. visual similarity), leading to a mismatch between the metaphor-like overextensions of children and the association-based metonymic extensions that dominate in language change and colexification. A more child-friendly version of our experiment could be used to test for a preference for similarity-based extension in children; more generally, our paradigm offers experimental control over the associative and similarity-based grounds for extension, potentially allowing us to explore conditions under which one or the other is preferred.

Finally, we only investigate extension in dyadic interaction; a worthwhile next step would be to explore extension in larger groups (e.g. using methods similar to those reported in Raviv et al., 2019a, 2019b, where members of larger groups interact dyadically). We expect that the reliability of the shared basis for grounding a semantic extension will become increasingly important in this context, meaning that only the most transparent extensions conventionalise in larger groups.

5. Conclusion

The flexibility of human language stems from our use of signals to provide evidence to prompt interlocutors to reconstruct our intended meaning. This allows signals to be used creatively to elicit innovative meanings, motivated by processes of semantic extension like metonymy and metaphor. Repeated use of innovative constructions leads to their linguistic entrenchment and to increases in the complexity of the meanings which can be communicated, notably in the emergence of grammatical items through grammaticalisation. In this paper, we present a method for studying semantic extension under experimental conditions, using an experimental semiotics approach. We show that participants can creatively extend existing labels to communicate new meanings in various semantic domains, using shapes as labels to represent colours, and then extending their use metonymically, to refer to objects stereotypically characterised by the relevant colours, and metaphorically, to emotions linked to colours more indirectly. Our manipulation of shape–colour associations early in the experiment shows that a grounding in a salient shared set of meaning-label associations facilitates both initial semantic extension and subsequent chaining of semantic extensions; we also find that metaphorical extension is particularly challenging, presumably due to the less obvious basis for grounding an extension from colour to emotion in our paradigm, but that the existence of some plausibly accessible motivation for semantic extension may be more important than its specific nature.

CRediT authorship contribution statement

Kenny Smith: Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation. **Josephine Bowerman:** Writing – original draft, Methodology, Conceptualization. **Andrew D.M. Smith:** Writing – review & editing, Writing – original draft, Software, Formal analysis.

Acknowledgements

This research received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (Grant 681942, held by K. Smith). Following the policies of the University of Edinburgh, for the purpose of open access we have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Data availability

The paper contains links to data and analysis code in an OSF repository.

References

Barcelona, A. (2003). On the plausibility of claiming a métonymie motivation for conceptual metaphor. In A. Barcelona (Ed.), *Metaphor and metonymy at the crossroads* (pp. 31–58). De Gruyter, http://dx.doi.org/10.1515/9783110894677.31.
Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*.
Brochhagen, T., Boleda, G., Gualdoni, E., & Xu, Y. (2023). From language development to language evolution: A unified view of human lexical creativity. *Science*, *381*, 431–436.
Bybee, J., Perkins, R., & Pagliuca, W. (1994). *The evolution of grammar: Tense, aspect and modality in the languages of the world*. Chicago, IL: University of Chicago Press.
Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
Culbertson, J. (2012). Typological universals as reflections of biased learning: Evidence from artificial language learning. *Language and Linguistics Compass*, *6*, 310–329.
Culbertson, J., & Adger, D. (2014). Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences*, *111*, 5842–5847. http://dx.doi.org/10.1073/pnas.1320525111.
Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, *122*, 306–329. http://dx.doi.org/10.1016/j.cognition.2011.10.017.
Dancygier, B., & Sweetser, E. (2014). *Figurative language*. New York: Cambridge University Press.
de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*, 1–12.
Deutscher, G. (2005). *The unfolding of language*. New York, NY: Metropolitan Books.
Fehér, O., Ritt, N., & Smith, K. (2019). Asymmetric accommodation during interaction leads to the regularisation of linguistic variants. *Journal of Memory and Language*, *109*, Article 104036.

Galantucci, B. (2009). Experimental semiotics: A new approach for studying communication as a form of joint action. *Topics in Cognitive Science*, *1*, 393–410.

Givón, T. (1979). *On understanding grammar*. New York, NY: Academic Press.

Goossens, L. (1990). Metaphtonymy: The interaction of metaphor and metonymy in expressions for linguistic action. *Cognitive Linguistics*, *1*, 323–342. http://dx.doi.org/10.1515/cogl.1990.1.3.323.

Hamilton, R. L. (2016). *Colour in English: from metonymy to metaphor* (Ph.D. thesis), University of Glasgow.

Haspelmath, M. (1998). Does grammaticalization need reanalysis? *Studies in Language*, *22*, 315–351.

Heine, B. (1997). *Cognitive foundations of grammar*. Oxford: Oxford University Press.

Heine, B., Claudi, U., & Hünnemeyer, F. (1991). *Grammaticalization: A conceptual framework*. Chicago, IL: University of Chicago Press.

Heine, B., & Kuteva, T. (2007). *The genesis of grammar: A reconstruction*. Oxford: Oxford University Press.

Hoefler, S. (2009). *Modelling the role of pragmatic plasticity in the evolution of linguistic communication* (Ph.D. thesis), University of Edinburgh.

Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization*. Cambridge: Cambridge University Press.

Jäger, G., & Rosenbach, A. (2008). Priming and unidirectional language change. *Theoretical Linguistics*, *34*, http://dx.doi.org/10.1515/THLI.2008.008.

Jonauskaite, D., Parraga, C. A., Quiblier, M., & Mohr, C. (2020). Feeling blue or seeing red? Similar patterns of emotion associations with colour patches and colour terms. *I-Perception*, *11*, Article 204166952090248.

Jonauskaite, D., Wicker, J., Mohr, C., Dael, N., Havelka, J., Papadatou-Pastou, M., Zhang, M., & Oberfeld, D. (2019). A machine learning approach to quantify the specificity of colour–emotion associations and their cultural differences. *Royal Society Open Science*, *6*, Article 190741.

Kapron-King, A., Kirby, S., Trousdale, G., & Smith, K. (2025). Grammatical unidirectionality is not reflected in individual preferences when performing artificial semantic extension. http://dx.doi.org/10.31219/osf.io/c5zks, (submitted for publication).

Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, *141*, 87–102.

Klepousniotou, E., & Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *Journal of Neurolinguistics*, *20*, 1–24.

Klepousniotou, E., Titone, D., & Romero, C. (2008). Making sense of word senses: The comprehension of polysemy depends on sense overlap. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *34*, 1534–1543.

Kövecses, Z. (2010). *Metaphor: A practical introduction*. Oxford: Oxford University Press.

Kövecses, Z. (2013). The metaphor–metonymy relationship: Correlation metaphors are based on metonymy. *Metaphor and Symbol*, *28*, 75–88. http://dx.doi.org/10.1080/10926488.2013.768498.

Kövecses, Z. (2020). *Extended conceptual metaphor theory*. Cambridge: Cambridge University Press.

Kuteva, T. (2008). *Auxiliation: An enquiry into the nature of grammaticalization*. Oxford: Oxford Univ. Press.

Kuteva, T., Heine, B., Hong, B., Long, H., Narrog, H., & Ree, S. (2019). *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.

Lakoff, G., & Johnson, M. (2003). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Martin, A., & White, J. (2019). Vowel harmony and disharmony are not equivalent in learning. *Linguistic Inquiry*, 1–20. http://dx.doi.org/10.1162/ling_a_00375.

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*, 280–291.

R Core Team (2023). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, https://www.R-project.org/.

Ramiro, C., Srinivasan, M., Malt, B. C., & Xu, Y. (2018). Algorithms in the historical emergence of word senses. *Proceedings of the National Academy of Sciences, USA*, *115*, 2323–2328.

Raviv, L., Meyer, A., & Lev-Ari, S. (2019a). Compositional structure can emerge without generational transmission. *Cognition*, *182*, 151–164.

Raviv, L., Meyer, A., & Lev-Ari, S. (2019b). Larger communities create more systematic languages. *Proceedings of the Royal Society B: Biological Sciences*, *286*, Article 20191262.

Roberts, G., & Clark, R. (2020). Dispersion, communication, and alignment: An experimental study of the emergence of structure in combinatorial phonology. *Journal of Language Evolution*, *5*, 121–139. http://dx.doi.org/10.1093/jole/lzaa004.

Roberts, G., Lewandowski, J., & Galantucci, B. (2015). How communication changes when we cannot mime the world: Experimental evidence for the effect of iconicity on combinatoriality. *Cognition*, *141*, 52–66.

Scott-Phillips, T. C. (2014). *Speaking our minds*. London: Palgrave Macmillan.

Scott-Phillips, T. C., Kirby, S., & Ritchie, G. R. (2009). Signalling signalhood and the emergence of communication. *Cognition*, *113*, 226–233.

Sievers, B., Lee, C., Haslett, W., & Wheatley, T. (2019). A multi-sensory code for emotional arousal. *Proceedings of the Royal Society B: Biological Sciences*, *286*, Article 20190513.

Smith, K., Ashton, C., & Sims-Williams, H. (2023). The relationship between frequency and irregularity in the evolution of linguistic structure: An experimental study. In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th annual conference of the Cognitive Science Society* (pp. 851–857). Austin, TX: Cognitive Science Society.

Smith, A. D. M., & Hoefler, S. (2015). The pivotal role of metaphor in the evolution of human language. In J. E. Díaz-Vera (Ed.), *Metaphor and metonymy across time and cultures* (pp. 123–140). Amsterdam: De Gruyter.

Smith, A. D. M., & Hoefler, S. (2017). From metaphor to symbols and grammar: The cumulative cultural evolution of language. In C. Power, M. Finnegan, & H. Callan (Eds.), *Human origins: contributions from social anthropology* (pp. 153–179). Oxford: Berghahn.

Sperber, D., & Wilson, D. (1995). *Relevance : Communication and cognition*. Oxford: Blackwell.

Svorou, S. (1994). *The grammar of space*. Amsterdam: John Benjamins Publishing Company.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*, 675–691.

Traugott, E. C., & Dasher, R. B. (2001). *Regularity in semantic change*. Cambridge: Cambridge University Press.

Traugott, E. C., & König, E. (1991). The semantics-pragmatics of grammaticalization revisited. In E. C. Traugott, & B. Heine (Eds.), *Approaches to grammaticalization. Volume I. Theoretical and methodological issues* (pp. 189–218). Amsterdam: John Benjamins Publishing Company.

Venables, W. N., Ripley, B. D., & Venables, W. N. (2002). Modern applied statistics with S, *Statistics and computing*, (4th). New York: Springer.

Verhoef, T. (2012). The origins of duality of patterning in artificial whistled languages. *Language and Cognition*, *4*, 357–380.

Verhoef, T., Kirby, S., & de Boer, B. (2014). Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics*, *43*, 57–68.

Verhoef, T., Marghetis, T., Walker, E., & Coulson, S. (2024). Brain responses to a lab-evolved artificial language with space-time metaphors. *Cognition*, *246*, Article 105763. http://dx.doi.org/10.1016/j.cognition.2024.105763.

Verhoef, T., Walker, E., & Marghetis, T. (2016). Cognitive biases and social coordination in the emergence of temporal language. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th annual conference of the cognitive science society* (pp. 2615–2620). Austin, TX: Cognitive Science Society.

White, J. (2014). Evidence for a learning bias against saltatory phonological alternations. *Cognition*, *130*, 96–115. http://dx.doi.org/10.1016/j.cognition.2013.09.008.

Wickham, H. (2009). *ggplot2*. New York, NY: Springer.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian Inference. *Psychological Review*, *114*, 245–272.