DE GRUYTER MOUTON
DOI 10.1515/langcog-2012-0017 — Language and Cognition 2012; 4(4): 297–318

Gareth Roberts and Bruno Galantucci

The emergence of duality of patterning: Insights from the laboratory

Abstract: The concept of duality of patterning (henceforth DP) has recently begun to undergo new scrutiny. In particular, the fact that Al-Sayyid Bedouin Sign Language (ABSL) does not appear to exhibit a layer of meaningless units (Sandler et al. 2011) casts doubt on the universality of DP as a defining feature of natural language. Why, then, do the vast majority of the world's languages exhibit DP? Two hypotheses have been suggested. The first is that DP is a necessary solution to the problem of conveying a large number of meanings; the second is that DP arises as a consequence of conventionalization. We tested these hypotheses in an experimental-semiotics study. Our results supported the hypothesis based on conventionalization but were inconclusive with regard to the hypothesis based on the number of meanings. At the same time, the task of measuring DP in an experimental-semiotics study presented interesting challenges, suggesting that the concept of DP may need some overhauling.

Keywords: dialogue, duality of patterning, emergence of phonology, experimental semiotics, human communication, sign language, social cognition

Corresponding author: Bruno Galantucci: Department of Psychology, Yeshiva University, 2495 Amsterdam Avenue, New York, NY, 10033, USA. E-mail: bruno.galantucci@yu.edu
Gareth Roberts: Yeshiva University.

1 Introduction

Since Hockett (1960) and Martinet (1960) introduced the concept of duality of patterning (henceforth DP) over half a century ago, it has generally been treated as a fundamental linguistic universal (e.g. Hurford 2002). DP is the phenomenon by which the utterances of a communication system can be analyzed as an arrangement of meaningless units (typically phonemes) into meaningful ones (typically morphemes). While the existence of the latter kind of unit rarely comes into question – any complex communication system must make use of meaningful units – the existence of a layer of meaningless structure, a property which we will here refer to as combinatoriality,[1] is less inevitable. Indeed, the question of whether or not a language exhibits DP typically boils down to the question of whether or not it exhibits combinatoriality (cf. Sandler et al. 2011: 504).[2] For known spoken languages, the answer to this question has been a resounding "yes". For sign languages too, the answer has overwhelmingly been "yes" ever since Stokoe's (1960) analysis of American Sign Language (Marshall 2011). Yet there are flies in the ointment. First, phonology itself and the nature of the phoneme as a unit of analysis have been undergoing increasing scrutiny (see e.g. Blevins 2012; Hayes 1995; Mielke 2008; Port 2010). Second, it appears that combinatoriality is neither entirely universal nor a necessary condition for a language to function. In fact, Sandler et al. (2011) have shown that Al-Sayyid Bedouin Sign Language (ABSL), despite being a fully-fledged functioning language with syntactic, morphological, and prosodic regularities, lacks combinatoriality below the phrasal level, as evidenced by a lack of minimal pairs, widespread violation of formal constraints on signs, and wide variation between signers in how they produce the same sign. Sandler et al. (2011) provide some evidence that combinatoriality may be starting to emerge in parts of ABSL; however, if combinatoriality can emerge in parts of a language, then this raises other issues. Most obviously, it raises the possibility that DP could be a feature of part of a system, not the whole. Second, and more fundamentally, it raises the question of how combinatoriality comes about and why it is so widespread, given that it characterizes pretty much the totality of known natural languages.

2 Explanations for the emergence of DP

Two factors in particular have been suggested as explaining the emergence of combinatoriality (and thus DP). The first is number of signs (henceforth set-size). Hockett (1960) noted that if a sign system based on sounds depends on analogue contrasts, it will become harder to distinguish signs auditorily as they increase in

---

[1] The term "combinatoriality" has been used in the literature in both a broad sense, which includes meaningful recombination (e.g. Jackendoff and Pinker 2005), and a narrow sense, to refer to meaningless recombination alone (e.g. Galantucci et al. 2010; Zuidema and de Boer 2009). We intend it in the latter sense and use it in preference to "phonology", since we do not wish to restrict ourselves to spoken language alone.

[2] While this is true of language and most other communication systems, it is not true of other complex human forms of expression such as music, which under normal circumstances can be analyzed as an arrangement of meaningless, but not meaningful, units. Cornish et al. (2010) and Verhoef et al. (2011), moreover, demonstrated the emergence of structure in entirely meaningless sequences in the laboratory.

number, creating a pressure to restructure the sound system in terms of discrete units (a mathematical model by Nowak et al. 1999, lent support to this hypothesis). That set-size is important for combinatoriality has also been suggested in production-based accounts of the origin of combinatoriality (Studdert-Kennedy 2000). Sandler et al. (2011: 526–536), by contrast, attribute the first manifestations of emerging combinatoriality in ABSL to conventionalization. Specifically, once an iconic sign becomes conventional, its iconicity may become dormant, leading to a loss of 'transparency' and a reinterpretation of the sub-elements of the sign as non-iconic or even meaningless. The sign for LEMON in ABSL, for instance, involves miming the act of squeezing a lemon. Since there is more than one way to squeeze a lemon, the form of this sign varies among signers (Sandler et al. 2011: 519), and mutual understanding is maintained via the link between the sign and the conceptual target, so long as this link remains evident to signers. A signer to whom it is not evident is likely to *reinterpret* a sign as arbitrary. Once such a reinterpretation has occurred, mutual understanding can be maintained only through increased regularity in the form of the sign,[3] which entails abstracting away from minor variation between instances of the same sign, which further encourages reanalysis in terms of discrete categories. The less transparent the sign-referent relationship, the greater the chance that reinterpretation will occur. This account, it is worth noting, relies on similar cognitive mechanisms to Hoefler and Smith's (2009) account of grammaticalization and can be seen as a special case of more general patterns of grammatical change.

The two explanations for the emergence of combinatoriality illustrated above are not mutually exclusive. Indeed, it is entirely possible that the emergence of combinatoriality is related to both set-size and transparency and that there is a complex relationship between the three. Since signs are easier to establish if there is greater opportunity for grounding them in something familiar (Galantucci 2005; Scott-Phillips et al. 2009), and transparent signs are by definition grounded, greater transparency should allow rapid growth in set-size, which – as suggested above – may in turn encourage greater combinatoriality. On the other hand, Sandler et al. (2011) explain combinatoriality as a response to *low* transparency. In other words, there is reason to expect both low transparency and high transparency to lead ultimately to combinatoriality, albeit by different routes. This may go some way to explaining the ubiquity of combinatorial structure in the world's languages. Moreover, given that the route from high transparency to combinatoriality is more indirect, it seems likely that combinatoriality takes

---

[3] Here, and elsewhere in the paper, we use the term "sign" not quite in the Saussurian sense of a signifier-signified pair (de Saussure 1916/1998), but in a sense more familiar from discussions of sign language: of a visual (or auditory) pattern used to represent some referent.

longer to arise in systems that afford highly transparent signs. This is consistent with the evidence from ABSL.

3 Testing explanations in the laboratory

Sandler et al.'s study of ABSL is a good example of the importance of recording and analyzing previously uninvestigated languages, particularly those that are very young or that exist in unusual cultural settings (cf. Everett 2005). By definition, however, unusual cultural settings are rare and hard to generalize from. Nor do brand new languages arise often, a fact that puts further severe limits on the extent to which generalizations can be made in this domain. If, moreover, we wish not only to demonstrate that some languages lack DP, but also to genuinely *explain* why they lack it, then studies with higher degrees of control are necessary. A particularly promising source of insight is to be found in the experimental-semiotics approach, that is, an approach that involves studying the emergence of novel human communication systems under laboratory conditions (Galantucci 2009; Galantucci et al. In press).

In the study we present here we have adopted such an approach to shed a preliminary light on the relationship between set-size, transparency, and combinatoriality in novel human communication systems.

3.1 Method

3.1.1 Participants

12 pairs of participants (4 female-female; 4 male-male; 4 mixed) participated in the study for course credit or monetary compensation.

3.1.2 The game

Participants played a cooperative guessing game, sitting in separate locations with the same set of four images (henceforth referents) displayed in random locations in a 5-by-5 grid on a video monitor (see Figure 1). The game consisted of a series of rounds. In each round, one player would play as "sender" and the other as "receiver". The sender was informed of a target referent and had to convey information to the receiver that would help the receiver select the correct target referent from among the referents visible to

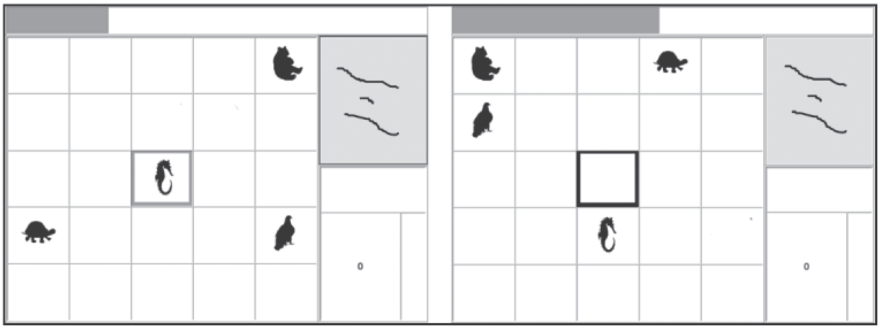

Fig. 1: Screenshot from early stage of game. The screen on the left was the Sender's screen; the screen on the right was the Receiver's.

them.[4] If the receiver selected the correct target the round was counted as successful; if not, the round was counted as unsuccessful. Since the players played in separate locations over the internet, they could not speak to each other directly. Instead, the sender could convey information to the receiver exclusively through the use of a digitizing pad and a magnetic stylus. The tracings that the sender made on this pad were transformed in a systematic way: While the horizontal component of the tracings determined the horizontal component of the output seen on the screen, the vertical component of the tracing did not determine the vertical component of the output. Rather, the vertical component of the output involved a simple downward movement at a constant rate (Figure 2A). These transformed tracings were relayed to the screens of both players in real time. Players could not use this pad as an effective drawing or writing device (Figure 2B), even after prolonged practice, and to succeed at the task pairs of players had to cooperatively develop novel forms of communication (Galantucci 2005). To help them in this, both players received feedback after each selection. Specifically, the receiver was shown what the target image had been and the sender was shown which image the receiver had selected. After the feedback phase, the next round began. Players swapped sender and receiver roles after each round.

The referents were presented as targets in a pseudo-random order: Pairs iterated through four referents twice every eight rounds (in random order). A performance score was kept updated for each referent, based on the proportion of successful rounds in the cycle. If a pair had at least 75% success on each of the

---

[4] To prevent participants from communicating the position of the referent rather than the referent itself, the referents were shown in locations on the grid which varied randomly for each round of the game and were different for the two players.

Fig. 2: (A) How the drawings players produced on the digitizing pad appeared on screen. (B) How common graphic symbols drawn on the digitizing pad appeared on the screen (adapted from Galantucci et al. 2010).

four referents, the number of referents in the set was increased to eight, and the cycle length was increased accordingly to 16 rounds. The referent set and cycle length continued to be incremented in this way until either players had mastered a set of 20 referents or two hours of playing had elapsed.

3.1.3 Referents

The referents used were black silhouettes of animals (see Figure 3). These silhouettes afforded the opportunity to develop signs with some degree of transparency, in which, for example, features of the silhouette (e.g. the trunk of the elephant) could be represented by a feature of the sign (e.g. a long curved line). However, the way in which their tracings were transformed did not allow players to reproduce the animal silhouettes or even to create simple drawings. In terms of the hypotheses described above, in other words, it was biased towards relatively low-transparency signs.

3.2 Results

3.2.1 Measures

All of the events in the game were recorded and three measures were derived from this data set: Set-size, Transparency, and Combinatoriality (see Table 1).

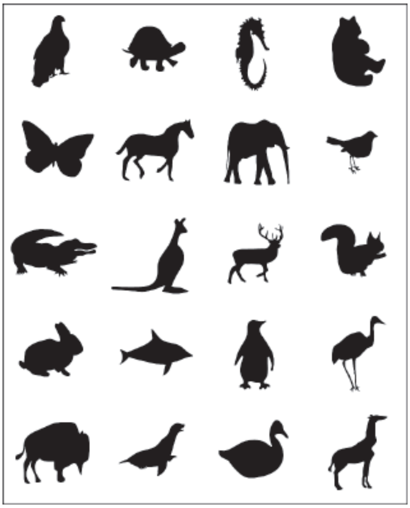

Fig. 3: Referents used in the game. The top row shows the referents that were visible to players at the start of the game. After they had reached 75% success on these four, the next row of referents was added, and so on.

Pair	Mean Set-size	Mean Combinatoriality	Mean Transparency (z-scores)
1	13.5	0.05	0.5
2	15.5	0.05	0.5
3	18	0.17	0.25
4	6	0.07	−0.25
5	12	0.02	0.75
6	11	0.01	0.25
7	15	0.06	0.5
8	20	0.05	1.75
9	18.5	0.05	2.5
10	16	0.11	0.25
11	15	0.05	1.35
12	15.5	0.09	0.5

Table 1: Summary of data

3.2.2 Set-size

For every referent that reached at least 75% success, the last successful sign produced by a given player for that referent was considered to be part of that player's communication system (henceforth player sign-set). The number of signs in this system thus provided a measure both of the Set-size and of the player's success at communicating. The Set-size for a pair was computed as the mean of the Set-sizes for the two players in the pair. The mean Set-size for the 12 pairs was 14.67 (SD = 3.75); the smallest sign-set contained six signs, and the largest contained 20.

3.2.3 Transparency

The more transparent the relationship between a sign and a referent, the easier it should be for an independent judge to match them up. Four judges, who had no previous familiarity with the signs or with the purpose of the study, matched signs with referents. This was done as follows.

First, the judges gained an understanding of the game by playing a few rounds themselves (as both sender and receiver, with pictures of faces as referents). Then they were shown a display containing one player's signs along with the referents they referred to; their task was to match the former with the latter (see Figure 4), taking as long as they wished. Once they had finished, another player's signs would appear. (The order in which the sign-sets appeared was randomized.) Each judge evaluated one player sign-set from every pair (12 sets in total) and every sign-set was shown to two judges.

The number of correct matches made by each judge for each player sign-set provides an index of the set's Transparency to that judge. However, this number cannot be used in its raw form because it is dependent on the size of the sign-set. In consequence, we converted it to a z-score by subtracting the mean number of correct matches we would expect, for that size of set, by chance (this was calculated using a Monte-Carlo simulation) and dividing the result by the standard deviation of that mean. Since every player sign-set was rated by two judges, the mean of the z-scores for the two judges was taken as the Transparency index for the set in question. Finally, the Transparency for a pair sign-set was computed as the mean of the Transparency for the two players in the pair. The overall mean Transparency for the 12 pairs was .73 (SD = .76), ranging from –.25 to 2.5.

3.2.4 Combinatoriality

Combinatoriality was measured using a slightly modified version of the Form Recombination Index used by Galantucci et al. (2010), which we briefly summarize here (see Appendix for a detailed description). This measure breaks a given

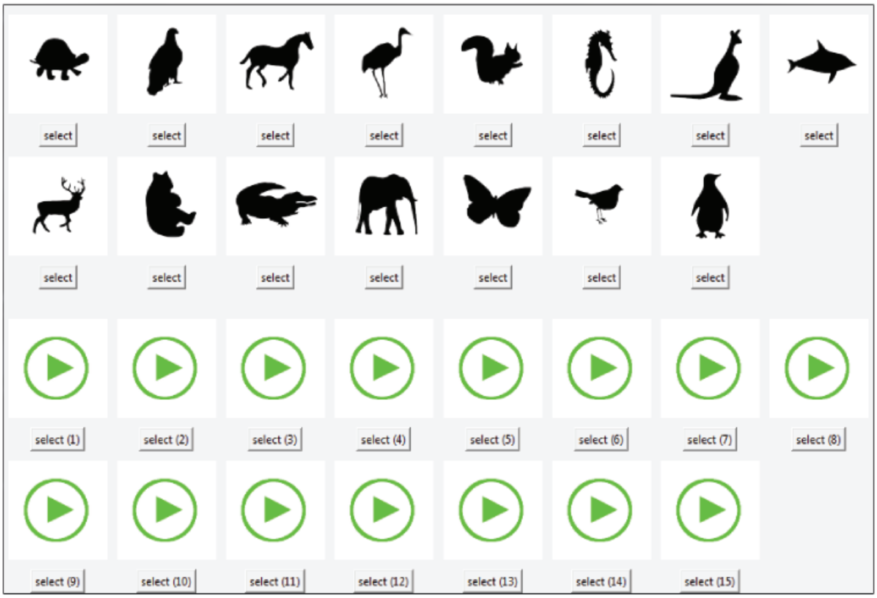

Fig. 4: Matching window. A judge could play a movie of a particular sign by clicking on a play icon, and match that sign with a referent by clicking the select button beneath it followed by the select button beneath the relevant referent (or vice versa). Matched referents were highlighted, and judges could change their minds as often as they liked.

sign into forms (parts of a sign divided by empty space). Forms within the sign are then compared with each other to remove duplicates, and the remaining forms are compared with all other forms in the system. The number of matches among these forms is then divided by the total number of comparisons to produce an index ranging from 0 to 1 (where 0 corresponds to a complete absence of Combinatoriality and 1 corresponds to maximal Combinatoriality). A system in which a small number of unique forms are reused many times will have higher Combinatoriality than a system in which a large number of forms are reused little. The mean Combinatoriality for the 12 pairs was .06 (SD = .04), ranging from .01 to .17.

3.2.5 Correlations

As can be seen in Figure 5, there was a strong positive correlation between Set-size and Transparency (r[10] = .65, p = .02), a positive correlation between Set-size

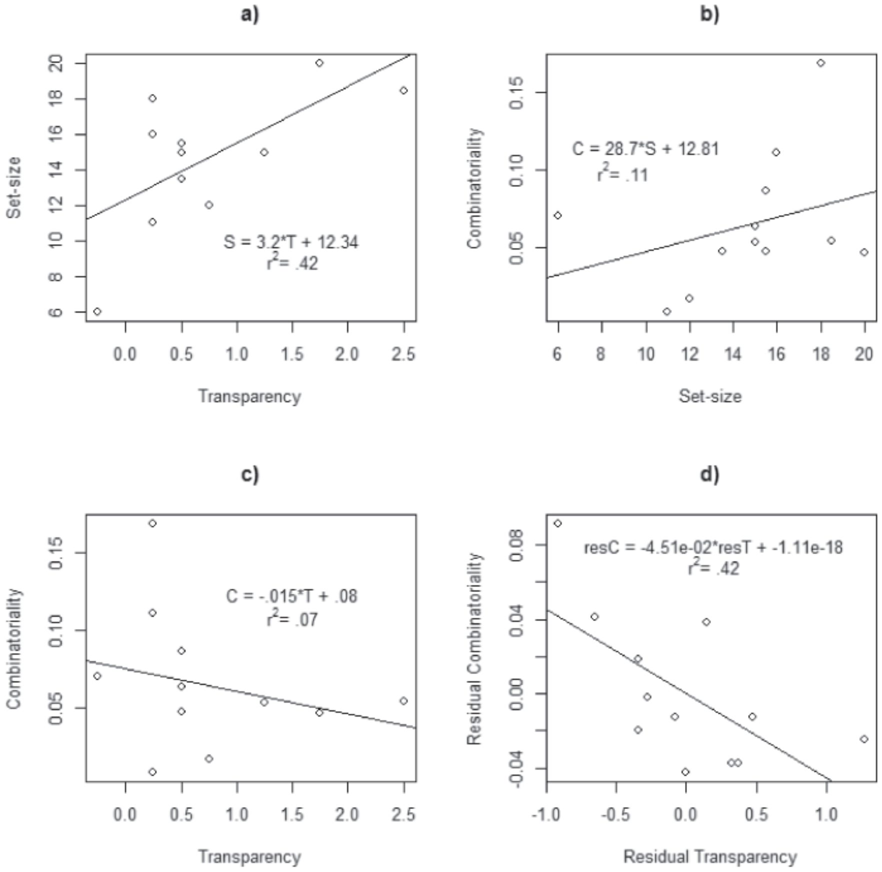

Fig. 5: Correlations between: a) Transparency (T) and Set-size (S); b) Combinatoriality (C) and Set-size; c) Transparency and Combinatoriality; d) Transparency and Combinatoriality, with Set-size partialed out.

and Combinatoriality (r[10] = .33, p = .3), and a negative correlation between Transparency and Combinatoriality (r[10] = −.26, p = .42). The strong correlation between Set-size and Transparency supports the hypothesis suggested above that more transparent signs are easier to ground, leading sign systems to grow faster. The presence of this correlation, however, poses a problem for interpreting the correlation between Transparency and Combinatoriality. That is, the positive correlation between Set-size and Combinatoriality interferes – via the positive correlation between Set-size and Transparency – with the negative correlation between Transparency and Combinatoriality (see Figure 6).

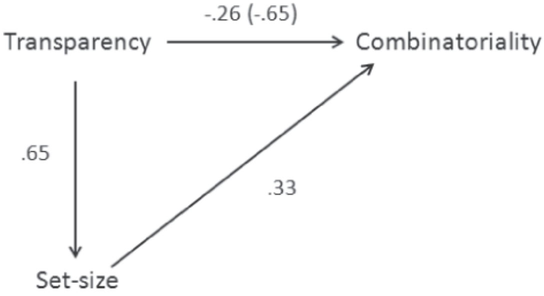

Fig. 6: Illustration of the relationship between Transparency, Combinatoriality and Set-size, with the correlation coefficients from our analysis. Arrows indicate the hypothesized direction of causality.

We therefore partialed out Set-size from the latter, and this revealed a much stronger correlation ($r[9] = -.65$, $p = .01$). This result is consistent with the hypothesis that Combinatoriality emerges as a response to low Transparency. The general pattern of results is also consistent with the hypothesis that high Transparency leads to Combinatoriality via Set-size (Figure 6), but the correlation between Set-size and Combinatoriality is too weak to say anything conclusive in this regard.

4 Discussion

The analysis presented above sheds light on the processes that affect the appearance of combinatoriality in communication systems. The strong negative correlation between Transparency and Combinatoriality, once Set-size was partialed out, lends support to Sandler et al.'s (2011) hypothesis that combinatoriality arises as a response to conventionalization. Overall, the pattern of results suggests the complex relationship illustrated in Figure 6 in which two pathways exist to higher Combinatoriality. The first leads directly from low Transparency; the second leads indirectly from high Transparency, through Set-size. However, this interpretation is limited by the weakness of the correlation between Set-size and Combinatoriality, which may be due to the small scale of our sign-sets: No set contained more than 20 signs. In general it must be stressed that there is a great difference in scale between natural languages and the communication systems we studied. ABSL, for instance, has existed for several generations, is used by over 150 people, and employs a great number of signs to communicate an infinite variety of meanings. The sign-sets analyzed above were developed over two hours by two individuals each and never consisted of more than 20 signs, which were used to communicate a finite set of meanings.

Aside from scale, there are other differences between our laboratory sign-sets and ABSL. It was noted above, for example, that the medium of communication encouraged relatively low-transparency signs (see Section 3.1.3.). Furthermore, while Sandler et al. (2011) were able to draw on research into other established sign languages to highlight the widespread violation in ABSL of formal constraints observed in these languages, a similar analysis could not be carried out on our data, since it is not clear what formal constraints might exist analogous to those found in natural languages (though see Section 4.3. below). Such differences, however, make all the more striking the similarities we found between features of the laboratory sign-sets and features of ABSL. In particular, Sandler et al. listed two further types of evidence for the lack of combinatoriality in ABSL, which we found in our own data: a dearth of minimal pairs, and wide variation between different speakers in how they produce the same sign.

Although there are a few cases of minimal pairs in our data (see Figure 10 below), the vast majority of these cases appear to be iconically motivated. In Pair 9, for example, the sign for a deer was distinguished from the sign for a horse by a form representing antlers (Figure 7).

There were also cases of variation between players in how they produced the same sign. Indeed, in only 30% of cases was a player's sign sufficiently close to his or her partner's sign to be counted as equivalent by the equivalence measure used in the FRI algorithm (see Appendix). It should be noted that this was not because pairs of players were using entirely different signs for the same referent. It was more as if their signs aimed at a common target, but which was underspecified in important respects. An alligator, for example, might be represented by both players using a jagged line, but with a different number of

Fig. 7: Two signs from the same player (Pair 9). The sign on the left represents a horse; the sign on the right represents a deer.

kinks in each case. In several cases, a player would produce the same shape as his or her partner, but mirrored with respect to the vertical axis. This is reminiscent of a comment by Sandler et al. (2011: 517) that “it is as if the signers are aiming for an iconic and holistic prototype, with details of formation taking a back seat.”

4.1 Measuring DP in the laboratory

The analysis presented above suggests that experimental-semiotics studies have the potential to make great contributions to investigating design features like DP. At the same time, however, it points at potential difficulties in doing so, particularly with regard to measurement. The measures used in the analysis described above are relatively conservative ones, which are sufficient to address the question we asked. However, they are too coarse-grained to answer certain questions relevant to investigating DP in the laboratory. In particular, they cannot reliably identify all the units of structure in a laboratory sign-set, or distinguish reliably between those units that are meaningless and those units that are meaningful. Both tasks turn out to be challenging, and in the following sections we discuss why this is the case.

4.1.1 Identifying units

The FRI algorithm involves dividing up a given sign into units. This division is conservative, since it identifies only one kind of unit: the pen-stroke. This makes it unlikely to identify very many spurious units, but it will miss any unit that is not separated from other units by a space (in other words, it would identify typed, but not cursive, letters). Any tool designed for the task is likely to be similarly limited: Since players cannot be expected to follow any single principle in segmenting their signs, there can be no principled way to be certain of having identified every unit in the data. Matters are more difficult still in the case of the referent-space. Segmentation of the sign-space must at least be grounded in the physical properties of the signs. This is not true of the referent-space, and players may use an infinite number of features to divide this up conceptually. Some of these divisions may be relatively straightforward, such as the division between bipeds and quadrupeds, but some may be very idiosyncratic. A player might, for instance, divide referents into those that remind him of his Aunt Vera and those that do not. This leads us to the problem of identifying meaningfulness.

4.1.2 Identifying meaningfulness

A system can be said to exhibit DP only if its signs are made up of subunits that are genuinely meaningless. For this reason it is useful to have a measure of meaningfulness. We define meaning as a reliable mapping between a unit in the sign-space and a unit in the referent-space in the mind of an individual. Researchers do not have direct access to that mind, and if a particular stroke in a player's communication system reliably maps to referents that remind him of his Aunt Vera, then there is a meaningful relationship that is very hard for the experimenter to detect.[5] It will also be hard for the player's partner to detect, of course, and it is worth noting that a pair of players may use the same signs for the same referents, but may not perceive them as being composed of the same units. Indeed, a participant who perceives an originally meaningful subunit as meaningless is more likely to reuse that subunit in an unrelated sign. This reuse may contribute to an increase in combinatoriality. This is the reverse of the behavior observed by Kirby et al. (2008), whose participants reanalyzed random correspondences as meaningful ones.

For the experimenter faced with the challenge of identifying meaningfulness, the nearest proxy is to identify reliable co-occurrence between units in the sign-space and units in the referent-space. In the analysis described in this paper, this is done most obviously at the level of whole referents and whole signs, both well-defined entities whose co-occurrence is easily measured. The 75% performance score threshold establishes the reliability of the mapping between them.[6] Below the referent-sign level, things get more complicated. The FRI identifies recombination, for instance, but is blind to the meaningfulness of the units that are recombined. This is not a relevant limitation on its use as a measure of combinatoriality. Communication systems are more wasteful with their meaningful units than with their meaningless ones: There are considerably more of them, and they are recombined less. FRI should therefore strongly correlate with true combinatoriality. The algorithm cannot be used, however, to identify the meaningfulness of an individual stroke.

The Transparency measure also does not measure meaningfulness, but identifies those sign-referent mappings that fall towards one end of a continuum. At

---

[5] The experimenter can, of course, ask the participant questions. However, it is often very difficult to phrase the question such that the participant answers usefully and in an unbiased way. See below for further discussion.

[6] Strictly speaking it is a measure of how reliably players understand each other, and in principle a player could successfully communicate the same referent using a different sign every time. This would be a very unusual strategy, however.

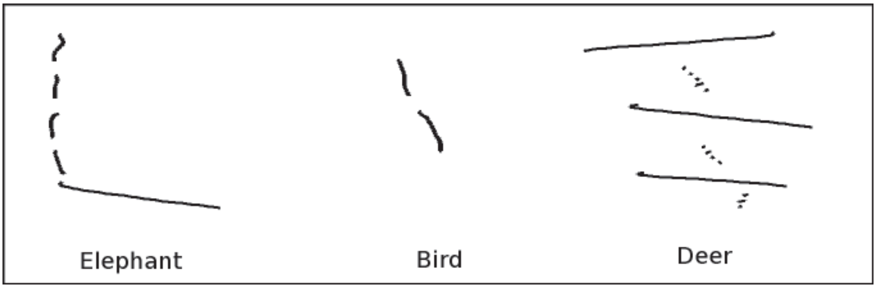

Fig. 8: Three signs from the same player in Pair 10. In the sign for the elephant, four short lines represent legs; similar lines apparently represent the bird's two legs; they are not used, however, in the sign for deer, an animal with four legs.

the extreme end of this continuum are sign-referent pairings (that do not and could not occur in the data) in which the sign is so iconic as to be identical to the referent. At the other end are elements in the sign-space and elements in the referent-space that co-occur reliably, but have no relationship with each other beyond their co-occurrence. Between these two extremes there are several other possibilities, all of which can be found in real-world languages:

a. A meaningful unit may co-occur with an apparently random subset of relevant referents; Figure 8 gives an example from our data in which the same stroke represents the legs of an elephant and the legs of a bird, but is not used in the sign for a deer. Similarly, the morpheme -*s* reliably co-occurs with plural meaning in English, but is not present in several plural words, such as *feet* and *children*.
b. A form that regularly co-occurs with a particular element in the meaning space may also appear in signs for entirely unrelated referents. For example, a jagged line might represent an alligator's teeth, a kangaroo's springiness, or a buffalo's fur (see Figure 9) Similarly in English, the *in-* prefix of *inflammable* does not imply negation, as it does in *insecure* or *inoperable*.
c. There may be a history effect such that sign-referent pairs established earlier share formal units with each other, but not with later pairs. English words of Anglo-Saxon origin, for example, are likely to lack certain sound combinations – e.g. /ski/ – that are present in words that were coined or borrowed after the early Anglo-Saxon period. This phenomenon may explain why Pair 10 used similar short lines to represent both the legs of a bird and the legs of an elephant (see Figure 8), but not to represent the legs of a deer: The elephant and the bird referents both appeared at the same time, earlier in the game than the deer referent (see Figure 3).

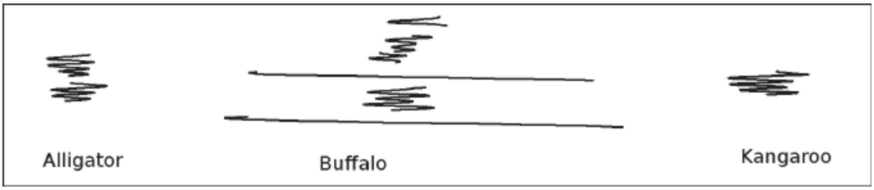

Fig. 9: A similar form appears in three signs by the same player (Pair 3) apparently representing three different features: an alligator's teeth, a buffalo's fur, and a kangaroo's springiness.

All of these cases cause problems of analysis in small datasets (including all laboratory sign-sets collected to date). Words like *men*, *women*, and *children* pose no problem to identifying plural morphemes in English. If in a laboratory sign-set, on the other hand, there were only five signs for plural referents, three of which had an element in common, it would be hard to tell whether the element should be treated as meaningful or meaningless, particularly if that element happened to reoccur in signs for non-plural referents, as indeed the *-s* suffix does in English.

4.2 The laboratory and the world

In more than one respect, DP may be easier to identify in natural languages than in the laboratory. First, as noted above, sign-sets are larger in real-world languages, making it easier to identify structural patterns. It is easier, for instance, to find minimal pairs in large combinatorial sign-sets than in small ones simply because there are more opportunities for them to arise. Second, a researcher investigating real-world languages usually has more speakers to call upon for information and has a greater opportunity to ask them follow-up questions days or even weeks after the original data were gathered. Several days after the experiment, participants are likely to have forgotten many details of the language they developed, and the experimenter must therefore ask all necessary questions as soon after the experiment as possible. In several respects the position of an experimenter looking for evidence of DP in a laboratory sign-set is similar to that of a researcher looking for evidence of DP in fragments of writing in a long-dead language.

However, it should not be interpreted from this that it is straightforward to identify the presence or absence of DP in a real-world language with a sizeable community of speakers. Identifying units and meaningfulness may typically be more difficult in the case of small laboratory sign-sets, but neither task is trivial for any language. Indeed, a close examination of DP in natural language reveals many complications (Ladd 2012) which, taken together with the difficulties encountered in the laboratory, suggest that DP is not quite carving nature at its joints. There are two ways in which this is the case. On the one hand, DP is clearly not monolithic, but a confluence of at least two independent phenomena (cf. Hockett 1961: 45): the decomposition of a system into discrete meaningful units, and the decomposition of those units into meaningless ones. Indeed, even the dependency implied by that wording should not be taken for granted. In principle meaningless units could arise first, and meaning attached to clusters of them. The other way in which DP does not quite carve nature at the joints is that it can be a feature not only of whole systems, but of parts of those systems. In what follows, we further elaborate on the latter issue.

4.3 DP as a non-system-wide phenomenon

It has traditionally been assumed that the set of signs exhibiting DP should roughly correspond to the system as a whole (Hockett 1960). This assumption may reflect an inherited tendency in linguistics to view a language as “un système où tout se tient”, but it is not an unreasonable expectation. If combinatoriality arises as a response to set-size, as Hockett hypothesized, then it would be surprising if it did not quickly come to characterize the entire set (albeit with marginal exceptions that resist analysis as part of the system, such as English “tsk-tsk”). But, as the analysis presented in this paper suggests, this is only part of the story. If combinatoriality is a response to low transparency in particular subsets of the system, then we might expect there at least to be a period of time where it characterizes those subsets. It is important to keep in mind that DP can never be a feature of individual signs. By its nature, DP must involve sets of signs; but these sets need not be equal to the system as a whole. Indeed, the suggestion above implies that it is in principle possible for a single system to contain multiple combinatorial systems, which might or might not overlap. This would be evidenced by some units never occurring together in the same word (to an extent beyond what would be expected by chance). The sign-sets in the laboratory data are too small to provide any clear evidence of this, but there are examples of combinatoriality not extending to the set as a whole. The system produced by Pair 3 is a good example. Figure 10 shows eight signs produced by a player in this pair (see Figure 9 for others).

Two of these signs can be easily interpreted in terms of meaningful units: The pair of jagged lines likely represent the alligator’s teeth, while the sign for the

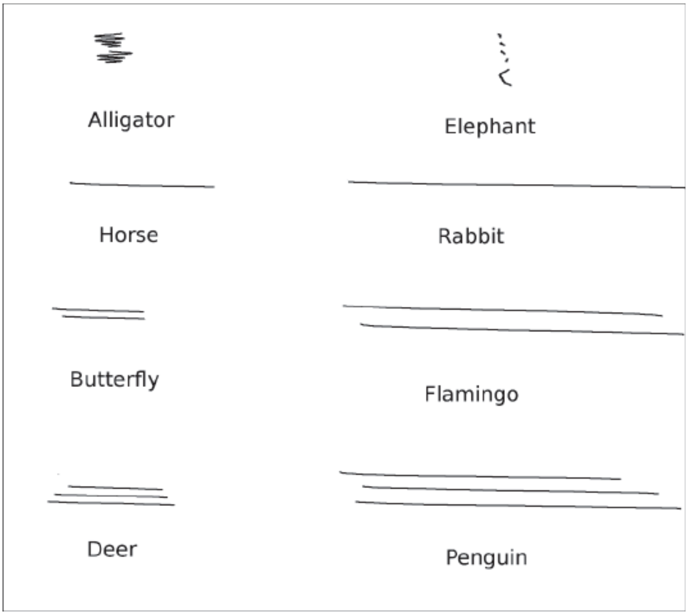

Fig. 10: Eight signs produced by one player in Pair 3. While the first two bear a relatively transparent relationship to their referent, the remaining six appear to be highly combinatorial.

elephant consists of five parts likely representing four legs and a trunk. The remaining six signs appear to exhibit combinatoriality and are represented by arrangements of either long or short horizontal lines. It is impossible to be certain that none of these signs are intended to bear an iconic relation to what they depict (see Section 4.1.2. above): The sign for the flamingo, for example, may be intended to represent the bird's legs; however, it is hard to perceive such a relation in the other cases. This set may even exhibit something analogous to a phonological constraint: Short and long lines are contrastive, but never appear in the same sign. In any case, it is hard not to see indications in these data of emerging combinatoriality, although it remains limited to a subset of signs. Of the 18 signs in this set, five – such as the alligator – seem clearly iconic in origin, while another three or four – such as the flamingo – are likely to be.

5 Conclusions

Over the last few years, the concept of DP has begun to undergo new scrutiny (Ladd 2012; Blevins 2012). In particular, the discovery of a natural language such as ABSL that does not appear to exhibit combinatoriality (Sandler et al. 2011) casts doubt on the universality of DP as a defining feature of natural language. But if DP is not a necessary feature of a complex communication system, why do the vast majority of the world's languages exhibit it? Theoretical work and the study of ABSL suggest two hypotheses to explain the emergence of DP. The first is that it arises as a solution to the problem of conveying a very large number of meanings (Hockett 1960; Nowak et al. 1999; Studdert-Kennedy 2000). The second hypothesis is that DP arises as a consequence of conventionalization, as iconic signs lose transparency (Sandler et al. 2011). However, the two hypotheses are not mutually exclusive, and while our analysis of laboratory data lends the most support to the second, our findings are consistent with the possibility that both low and high transparency lead to combinatoriality by different routes. In addition, the challenge of measuring DP in simple laboratory sign-sets reveals a number of subtle intricacies in the concept of DP, suggesting that it may need some overhauling.

References

Blevins, J. 2012. Duality of patterning: Absolute universal or statistical tendency? *Language and Cognition* 4(4).

Cornish, H., M. Christensen & S. Kirby. 2010. The emergence of structure from sequence memory constraints in cultural transmission. In A. D. M. Smith, M. Schouwstra, B. De Boer & K. Smith (eds.), *The evolution of language: Proceedings of the 8th international conference (Evolang 8)*, 387–388. Singapore: World Scientific.

de Saussure, F. 1916/1998. *Course in general linguistics*. (R. Harris, Trans. Reprint ed.). Chicago, IL: Open Court Publishing Co.

Everett, D. L. 2005. Cultural constraints on grammar and cognition in Pirahã – Another look at the design features of human language. *Current Anthropology* 46(4). 621–646.

Galantucci, B. 2005. An experimental study of the emergence of human communication systems. *Cognitive Science* 29(5). 737–767.

Galantucci, B. 2009. Experimental semiotics: A new approach for studying communication as a form of joint action. *Topics in Cognitive Science* 1(2). 393–410.

Galantucci, B., S. Garrod & G. Roberts. (2012) Experimental semiotics. *Language and Linguistic Compass*. 6(8). 477–493.

Galantucci, B., C. Kroos & T. Rhodes. 2010. The effects of rapidity of fading on communication systems. *Interaction Studies* 11(1), 100–111.

Hayes, B. 1995. On what to teach the undergraduates: Some changing orthodoxies in phonological theory. In I.-H. Lee (ed.), *Linguistics in the morning calm 3*, 59–77. Seoul: Hanshin.

Hockett, C. F. 1960. The origin of speech. *Scientific American* 203(3). 88–96.

Hockett, C. F. 1961. Linguistic elements and their relations. *Language* 37(1). 29–53.

Hoefler, S. H. & A. D. M. Smith. 2009. The pre-linguistic basis of grammaticalisation: A unified approach to metaphor and reanalysis. *Studies in Language* 33(4). 886–909.

Hurford, J. R. 2002. The role of expression and representation in language evolution. In A. Wray (ed.), *The transition to language*, 311–334. Oxford: Oxford University Press.

Jackendoff, R. & S. Pinker. 2005. The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser and Chomsky). *Cognition* 97(2). 211–225.

Kirby, S., H. Cornish & K. Smith. 2008. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences of the United States of America* 105(31). 10681–10686.

Ladd, D. R. 2012. What is duality of patterning, anyway? *Language and Cognition* 4(4).

Marshall, C. R. 2011. Sign language phonology. In N. C. Kula, B. Botma & K. Nasukawa (eds.), *The continuum companion to phonology*, 254–277. London: Continuum.

Martinet, A. 1960. *Elements of general linguistics* (E. Palmer, Trans.). Chicago: University of Chicago Press.

Mielke, J. 2008. *The emergence of distinctive features*. Oxford: Oxford University Press.

Nowak, M. A., D. C. Krakauer & A. Dress. 1999. An error limit for the evolution of language. *Proceedings of the Royal Society of London Series B-Biological Sciences* 266(1433). 2131–2136.

Nowak, M. A., J. B. Plotkin & D. C. Krakauer. 1999. The evolutionary language game. *Journal of Theoretical Biology* 200(2). 147–162.

Port, R. F. 2010. Language as a social institution: Why Phonemes and words do not live in the brain. *Ecological Psychology* 22(4). 304–326.

Sandler, W., M. Aronoff, I. Meir & C. Padden. 2011. The gradual emergence of phonological form in a new language. *Natural Language and Linguistic Theory* 29(2). 503–543.

Scott-Phillips, T. C., S. Kirby & G. R. S. Ritchie. 2009. Signalling signalhood and the emergence of communication. *Cognition* 113(2). 226–233.

Stokoe, W. C. 1960. Sign language structure: An outline of the visual communications systems. *Studies in Linguistics, Occasional Papers* 8. Buffalo, NY: University of Buffalo.

Studdert-Kennedy, M. 2000. Evolutionary implications of the particulate principle: Imitation and the dissociation of phonetic form from semantic function. In C. Knight, M. Studdert Kennedy & J. R. Hurford (eds.), *The evolutionary emergence of language*, 161–176. Cambridge: Cambridge University Press.

Verhoef, T., S. Kirby & C. Padden. 2011. Cultural emergence of combinatorial structure in an artificial whistled language. In L. Carlson, C. Hölscher & T. Shipley (eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, 483–488. Austin, TX: Cognitive Science Society.

Zuidema, W. & B. de Boer. 2009. The evolution of combinatorial phonology. *Journal of Phonetics* 37(2). 125–144.

Acknowledgements

We thank Christian Kroos, who authored the program that runs the game used for the study, Carrie Theisen, who helped us in refining the FRI algorithm, and Theo Rhodes, who helped us with the data collection. We also gratefully acknowledge the support of the National Science Foundation (BCS-1026943).

Appendix

Combinatoriality was measured using a slightly modified version of the Form Recombination Index used by Galantucci et al. (2010). It involved a series of steps:

Identifying forms: Every sign was broken down into *forms* – vectors of contiguous stylus positions (*sample points*, sampled at approximately 32 Hz) which corresponded on the digitizing pad to strokes separated by white space. Since participants might briefly raise the pen from the stylus by accident, gaps of three sample points of fewer were disregarded for this purpose.

Form equivalence test: Forms were compared using a procedure based on their shape. For each form, we computed its mean value and determined in how many places the form crossed the mean value (henceforth *mean-crossings*). Then, we computed the proportion of the form that fell between each mean crossing (henceforth *form proportions*; see Figure 11) and expressed the shape of the form as the ordered series of its form proportions. Two forms were considered

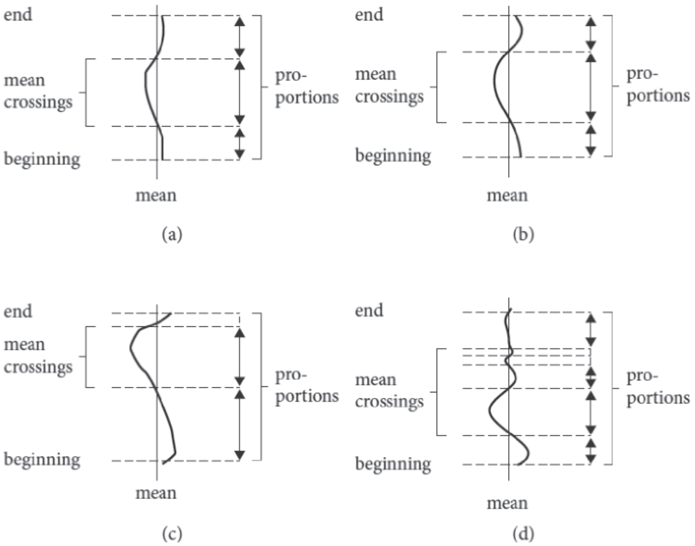

Fig. 11: The form equivalence test. The forms in (A) and (B) pass the test because they have the same number of mean crossings and their form proportions are within 10% of each other. The forms in (a) and (c), as well as the forms in (b) and (c), do not pass the test because their form proportions are not within 10% of each other. The form in (d) is not equivalent to any of the other forms because it has a different number of mean crossings (adapted from Galantucci et al. 2010).

equivalent if they had the same number of mean-crossings and their respective form proportions had values that were within ten percent of each other. Our measure differed from that described by Galantucci et al. (2010) in that we used the form equivalence test twice. First we used it *within* signs to identify unique forms: That is, if a sign contained more than one equivalent form, one of these was selected at random and the rest discounted. We then used the equivalence test again to compare unique forms between signs.

Form recombination index: Applying the form equivalence test to each pair of unique forms in the set resulted in a matrix which indicated, for each form, how many times that form recurred in the whole set. To obtain an index of form recombination for each player, we then calculated how frequently forms recurred across the database, and divided that by the total number of unique forms.