Foundations of Representation: Where Might Graphical Symbol Systems Come From?

Simon Garrod\textsuperscript{a}, Nicolas Fay\textsuperscript{b,c}, John Lee\textsuperscript{d}, Jon Oberlander\textsuperscript{d}, Tracy MacLeod\textsuperscript{a}

\textsuperscript{a}Department of Psychology, University of Glasgow, Scotland
\textsuperscript{b}Department of Psychology, University of Western Australia, Australia
\textsuperscript{c}ATR Media Information Science Labs, Japan
\textsuperscript{d}School of Informatics, University of Edinburgh, Scotland

Received 5 April 2006; received in revised form 23 November 2006; accepted 4 January 2007

Abstract

It has been suggested that iconic graphical signs evolve into symbolic graphical signs through repeated usage. This article reports a series of interactive graphical communication experiments using a ‘pictionary’ task to establish the conditions under which the evolution might occur. Experiment 1 rules out a simple repetition based account in favor of an account that requires feedback and interaction between communicators. Experiment 2 shows how the degree of interaction affects the evolution of signs according to a process of grounding. Experiment 3 confirms the prediction that those not involved directly in the interaction have trouble interpreting the graphical signs produced in Experiment 1. On the basis of these results, this article argues that icons evolve into symbols as a consequence of the systematic shift in the locus of information from the sign to the users’ memory of the sign’s usage supported by an interactive grounding process.

Keywords: Graphical representation; Symbols; Icons; Signs; Communication

Symbols grow. They come into being by development out of other signs, particularly from icons, or from mixed signs partaking of the nature of icons and symbols. (Peirce, 1931–1958, Vol. 2, p. 302)

1. Introduction

From the cave paintings of Lascaux to the earliest pictographic writing systems of the ancient Egyptians, early human settlements are marked by the use of simple graphical representations. Even today, graphical representation plays an important part in various sign systems, such as road traffic signs or the signs on a computer interface. In some cases they

Correspondence should be addressed to Simon Garrod, Department of Psychology, University of Glasgow, 56 Hillhead Street, Glasgow, G12 8QT, United Kingdom. E-mail: simon@psy.gla.ac.uk
serve as direct iconic representations of objects (e.g., the trash can icon on many computer interfaces), but in other cases they serve as more abstract symbolic representations (e.g., the red triangle motif used in traffic signs to indicate warning). Like Peirce, some argue that symbolic graphical representations evolved from earlier iconic representations. For example, the Assyrian symbolic writing system evolved from the iconic pictographic system of early Cuneiform via early Babylonian (Tversky, 1995). A similar observation can be made about the evolution of Chinese characters (see Fig. 1). It has also been suggested that modern symbolic traffic signs incorporating the red triangle motif evolved from an earlier traffic sign that indicated danger (Krampen, 1983). Yet, little is known about how or why icons evolve into symbols. This article addresses the question of how graphical signs can become symbolic by investigating the conditions under which graphical representations evolve during the course of a graphical communication task.

Consider the following experimental setup. Pairs of participants are asked to communicate concepts to each other using only graphical means. The setup is similar to the popular parlor game *Pictionary*, except that the concepts are drawn from a fixed list that is known to both participants. One participant (the matcher) has to identify the concept drawn on a whiteboard by the other participant (the director) from a set of 16 alternatives. These include groups of easily confused concepts such as art gallery, museum, parliament, theatre; Robert De Niro, Clint Eastwood, Arnold Schwarzenegger. The participants are able to make some minimal assumptions about each other’s shared cultural background (e.g., they might expect the other to know that Schwarzenegger has large muscles), but have relatively few shared resources for using the communication system. During the experiment, they carry out the task through a number of blocks, each consisting of 12 of the same 16 items, chosen arbitrarily, so the items regularly recur, although at random intervals.

We will show that when participants carry out this task under most but not all conditions, three things happen to their drawings as illustrated by the sequence in Fig. 2. As the game proceeds from block to block, the drawings change: (a) They become graphically simpler (e.g., they contain fewer lines, and fine details are lost); (b) they become more schematic or abstract as representations of the concepts they stand for; and (c) when directors alternate with each other, their drawings increasingly converge with those of their partner.

What proves to be crucial for the evolution of the graphical representations in these three ways is the degree to which the participants can interact graphically with each other during the experiment. Our basic argument will be that interactive graphical communication enables participants to develop symbolic representations from what started out as primarily iconic
representations through a “grounding process” similar to that proposed for interactive spoken communication (Clark & Brennan, 1991).

The article is organized into four further sections. In the first, we develop a framework for analyzing different kinds of graphical representation based on Peirce’s account of signs, using ideas from information theory. In the second, we use this framework to outline our main hypotheses: (a) that communicative interaction enables iconic graphical representations to evolve rapidly into symbolic representations and (b) that the evolution involves a shift in the locus of the information from the graphical sign to the mental representation required for successful interaction. In particular, we will argue that during the evolution from iconic to symbolic graphical representation, structural complexity migrates from the sign to memory representations in sign users. The third section of the article reports three graphical communication experiments that test these predictions. Finally, the fourth section considers the implications of the experimental results for both communication and the nature of graphical symbol systems.

2. Classifying representations as icons, indexes, and symbols

According to Peirce (1931–1958), a sign is a relation among three elements: the sign, the object being signified, and what he called the interpretant or understanding of the sign (see also Greenlee, 1973; Krampen, 1983). For example, the trash icon on a computer screen is a sign that relates to an object (i.e., the memory deletion buffer of the computer system) and has an interpretant (i.e., the understanding that it represents where to drag and drop another
For Peirce, it is the presence of all three elements that defines signification. But he distinguished between three kinds of sign, which he called icon, index, and symbol. The three kinds of sign differ in terms of how the sign relates to its object. For icons there must be a salient perceptual/structural resemblance between sign and object (e.g., the picture of the trash can in the computer icon resembles a trash can itself). Most pictorial representations are therefore icons, because they usually reflect the appearance of the signified object. By contrast, indexes relate to their objects in a causal way. For example, the level of mercury in a barometer is an index of atmospheric pressure; it does not resemble atmospheric pressure but it is causally linked to it. Similarly, smoke is an index of fire. Finally, symbols relate to their objects, neither in terms of resemblance nor causality, but purely on the basis of habit or convention. Thus, a symbol has to have been habitually or conventionally used to signify its object in order to operate as a sign for that object. Unlike icons and indexes, symbols only bear an arbitrary relation to their objects. For example, both the words dog and chien act as symbols for canine creatures, but there is neither a resemblance between the sounds of the words and the sounds made by canine creatures nor is there any relationship between the sounds in the two words themselves.

Peirce’s (1931–1958) distinction between three kinds of sign is intuitively appealing. However, when it comes to classifying graphical signs of the sort shown in Fig. 2, it often seems that the distinction between iconic and symbolic signs is not binary, but rather a matter of degree. For example, the depiction of ‘cartoon’ in Block 1 of Fig. 2 is clearly iconic because it resembles a cartoon character, as does the depiction in Block 4. However, the later depiction in Block 6 of Fig. 2 has become so schematic that it no longer resembles the cartoon character and so becomes more a symbol for ‘cartoon.’ In other words, its interpretation now depends to a greater degree on the fact that the part of the drawing (i.e., the part resembling the top of the cartoon character’s ears) has been habitually used to signify its object ‘cartoon’ over the course of the ‘pictionary’ task and so is now sufficient to bring the object to the interpreter’s mind. Nevertheless, the depiction still retains something of its original iconicity because it is still possible to see how it resembles a part of the cartoon character. So, like Krampen (1983) and to a certain extent Peirce (1931–1958) himself we treat the distinction between iconic and symbolic representations as graded. Similar arguments can be applied to the distinction between index and icon. For example, a photograph is usually taken to be an index of its subject, because the pattern of pixels on the photographic paper is caused by the light reflecting from the subject. However, the photo is also iconic to the extent that it resembles the subject. So, we shall treat signs as exhibiting different degrees of iconicity, symbolicity, and indexicality.

The difference between icons and symbols can be characterized in information theoretic terms by looking at how information is distributed between the sign itself and the user’s knowledge of how the sign has been used in the past (Shannon, 1948). Because iconic signs depend on a structure mapping between sign and object, they carry information in their graphical structure. If a sign for an object maps some of the structure of the object, then we can see that the structure in the sign is carrying information. And, this sign is functioning iconically. Other things being equal the more structured the sign, the more iconic the sign is. On the other hand, when symbols are used complexity (or information) resides in the users’ knowledge of previous use of the symbol. To this extent a symbol only requires enough
structure to be recognizable as that symbol distinct from any other. Hence, symbols can be graphically simpler than icons.

We can now express key differences between icons and symbols of the kind seen in the ‘pictionary’ task. Over an extended sequence of interactions, we hypothesize there is a shift of the locus of information from the sign itself to the communicators’ representations of the sign’s usage. At the beginning of the ‘pictionary’ experiment communicators need to be able to recognize the object from the sign and they can only do this through the visual resemblance between sign and object. So, the sign functions iconically. Sometime later in the interaction, the same concept needs to be communicated. Now, a much-reduced drawing serves to secure recognition. What differs is the shared experience of the interaction that the participants now possess. In particular, recognition of a sign will involve checking the current sign against a memory representation built up from a variety of previous recognition events, classified as involving variants of the same sign. To the extent that resemblance matters at all, it is resemblance between sign and sign, not between sign and object. Depending on the precise conditions of interaction, particular aspects or features of the original iconic drawing will have been accentuated and others diminished or removed. Under such circumstances the sign is functioning symbolically. The more structured the interaction history behind the sign, the more symbolic it can become.

Indexes should undergo a similar transformation. Whereas an icon visually resembles its object, an index is only causally related to its object; but there still has to be a visual representation of the index, and to begin with this will represent features of the index iconically. For example, to signify ‘fire’ with the index ‘smoke’ we need an iconic depiction of smoke. Via interaction, the iconic component of any indexical sign should become simpler and thereby increasingly symbolic. In what follows, we will rarely distinguish such index-based icons from icons in general.

To sum up, the Peircean notions of icon, symbol, and index are useful tools in approaching the topic of graphical representation. Thinking about where structure (and hence, information) lies within an interaction helps us see how signs can function in more than one way at once, and that a sign may function one way at one time, and another way later on. We have suggested that if an interaction is structured in some way, the burden of information-carrying will shift from individual signs into history. As a consequence, what began as a graphically complex icon can evolve into a graphically simpler and more abstract symbol. But, if this does happen in the way we have supposed, what mechanism underlies it? Which kinds of interaction encourage the evolution of icons and indexes into symbols? The experiments we report in this article directly address these questions.

3. Hypotheses about how iconic and indexical representations evolve into symbols

In the Introduction, we made some preliminary observations about how the drawings in the ‘pictionary’ task change when repeated over the course of the experiment. Basically, they seem to become graphically simpler and more schematic with repeated use, evolving from a clear iconic representation of their object to a more abstract symbolic representation.
What is responsible for this shift? We consider two hypotheses: (a) simplification and refinement through repeated production of the sign and (b) simplification and refinement through grounding.

According to the first hypothesis, the simplification of the drawings comes about through repeated production. The idea is that directors (i.e., those drawing the pictures) will increasingly refine their drawings on the basis of their prior experience and come to depict only the important aspects of the object in the drawing. This could occur via a process along the lines of Bartlett’s (1932) “method of serial reproduction” whereby with repeated retelling stories become increasingly schematic and stereotypical. In the ‘pictionary’ task, each time the director has to depict a concept that he has previously depicted he recalls only the most prominent (or effective) elements of the previous depiction. In this way, as the task proceeds the director will learn to produce the most efficient sign—the one that is most unique and diagnostic. Tversky (1995) offers a version of this hypothesis to explain the increased simplification in written signs as they evolve from pictographic to more abstract writing systems. Notice that for simplification-by-repetition, feedback from the interpreter of the sign is not required.

However, for simplification-by-grounding, feedback of some kind is required. According to this second hypothesis graphical simplification occurs because the interpreter can indicate when the drawing is sufficiently clear to identify the object and so ground that sign. Successful grounding then helps to fix the sign–object interpretation in both the director’s and the matcher’s memories. Notice that when there is no grounding or it is unsuccessful (e.g., when a matcher queries a drawing) then the sign–object interpretation is less likely to be fixed in memory. With repeated use of the sign (i.e., the drawing) increasingly less information will be needed to recover the sign–object interpretation. The drawer can then increasingly simplify the sign so long as it can be grounded on each occasion of use. To the extent that simplification through grounding depends upon feedback, we can further hypothesize that the rate and degree of simplification will depend upon the quality of the communicative feedback that is given. Notice that the grounding hypothesis is similar in certain respects to Brennan and Clark’s (1996) idea that conversationalists use feedback to ground conceptual pacts whereby they can subsequently refer to objects with idiosyncratic shorthand descriptions (e.g., ice skater to refer to a particular tangram figure). The difference is that the graphical simplification, unlike the verbal simplification, enables a change from a complex iconic representation to a simpler symbolic representation.

Opportunity for giving feedback can be manipulated in various ways in the ‘pictionary’ task. First, one can have situations in which there is a single director and a single matcher throughout the experiment, where the matcher is allowed to offer graphical feedback but not required to depict the object themselves. For example, the matcher can indicate when they have “understood” the drawing by producing a tick next to the drawing or can annotate the director’s drawing in an attempt to check their interpretation. Second, one can have a situation in which both communicators act as director or matcher at different points in the experiment. So, both at some time depict the same object. This means that a matcher who subsequently draws a concept in a similar fashion to that of the director can demonstrate his or her understanding of the drawing and both participants can increasingly reduce the information in their pictures on the basis of this reciprocal understanding of the sign–object interpretation. This process should eventually lead to sign–object interpretations that represent
what is mutually intelligible to the two participants. When both participants act as director we predict two outcomes: (a) Participants should become better at identifying the drawings, and (b) their respective drawings should start to converge. It is also possible to further manipulate the quality of the interactive feedback by allowing both director and matcher to be co-present during the drawing process as opposed to separating them during drawing. The experiments reported here manipulate feedback in all these ways.

4. Experiments

To test the two basic hypotheses, the first experiment compared graphical communication under three conditions. First, there was a Single Director with no Feedback (SD-F) condition in which one participant repeatedly drew items but for an imaginary audience. This tests whether repeated production alone is sufficient to produce refinement and simplification. Second, there was a Single Director with Feedback (SD+F) condition in which there were two participants present: one who drew (the director) and one who matched (the matcher). This condition allowed for simple feedback and so tests the basic grounding hypothesis. Finally, there was a Double Director with Feedback (DD+F) condition in which the role of director and matcher alternated between the two participants between blocks. This meant that both participants acted as directors and matchers over the course of the experiment. The DD+F condition tests the above predictions that reciprocal production should improve communicative success and that the two directors’ drawings should converge.

In the second experiment, we explore the Double Director situation further by varying the degree of interaction available to the two participants. We contrast a DD+F condition of Experiment 1 (non-concurrent feedback condition) with a condition in which both director and matcher are co-present during the task affording a higher degree of interaction (concurrent feedback condition). In the third experiment, we test whether non-participants (“overseers”) have difficulty in understanding graphical signs compared to the original participants.

4.1. Experiment 1

The experimental task was the graphical equivalent of a verbal referential communication task (Krauss & Weinheimer, 1964, 1966, 1967). Like the game Pictionary, participants were required to depict various concepts in such a way that a partner could identify them. Again, like Pictionary, participants were not allowed to speak or use letters in their drawings. Unlike Pictionary, concepts were drawn from a list of 16 items, which were known to both participants. Concepts included easily confusable items such as drama, soap opera, theatre, Clint Eastwood, and Robert De Niro (see Fig. 3 for a complete listing).

The director, or drawer, was instructed to depict the first 12 items from an ordered list (12 targets plus 4 distracters) such that the matcher could identify each drawing from their unordered list.
4.1.1. Design and procedure

The experiment had three conditions. In two conditions (SD-F, SD+F) there was a single director throughout the experiment, and each director was paired with one matcher. However, in the SD-F condition, the matcher only saw and identified the drawings once the director had finished drawing all the items over a series of four blocks, so there was no matcher feedback. In the SD+F condition, the director and matcher were separated by a screen while the drawing was taking place. However, after each drawing was completed, the matcher was taken around the screen to identify it. At this point, the matcher was able to give graphical feedback. Matchers were instructed that they could annotate or modify the director’s drawing or simply indicate that they had identified the concept by marking it with a tick. The director could then modify the drawing and if necessary receive further graphical feedback until the matcher was satisfied that he or she had interpreted the picture correctly. In the third DD+F condition participants alternated roles (director and matcher) from block to block. Otherwise, the instructions were identical to that in the SD+F condition. Contrasting task performance (i.e., accuracy of identification of the drawings by the matcher) in SD+F versus DD+F conditions tests whether reciprocal production and comprehension of drawings improves mutual understanding, as predicted above. It also makes it possible to investigate both graphical refinement and convergence in drawings from the two communicators (we leave analysis of convergence until after reporting Experiment 2).

Participants went through four blocks in the SD-F condition. In Block 1, the single directors drew 12 of the 16 target items in isolation. In Block 2, they again drew 12 of the 16 items (in a different random order) in isolation. They did this twice more (Blocks 3 & 4). Thus, each participant drew a total of 48 items, with each of the 12 target items recurring four times. Participants in the SD+F condition also drew 12 of the 16 items but across six blocks, so that each item recurred six times during the experiment. Unlike participants in the SD-F condition, after each drawing was completed, the matcher was allowed to give feedback (see above). In the DD+F condition, participants again drew 12 of the 16 items (presented in a different random order for each block) over six blocks, but in this condition the director and matcher alternated across blocks. So, by the end of the experiment each participant had drawn the 12 items three times as director and matched 12 drawings of the same items three times as matcher. Drawing took place on a standard whiteboard using whiteboard marker pens, with each participant using a different colored pen. The different colors were used to aid in the analysis of the feedback. The completed drawings were recorded on digital camera for subsequent analysis.

<table>
<thead>
<tr>
<th>Places</th>
<th>People</th>
<th>Programmes</th>
<th>Objects</th>
<th>Abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theatre</td>
<td>Robert De Niro</td>
<td>Drama</td>
<td>Television</td>
<td>Loud</td>
</tr>
<tr>
<td>Art gallery</td>
<td>Arnold Schwarzenegger</td>
<td>Soap opera</td>
<td>Computer monitor</td>
<td>Homesick</td>
</tr>
<tr>
<td>Museum</td>
<td>Clint Eastwood</td>
<td>Cartoon</td>
<td>Microwave</td>
<td>Poverty</td>
</tr>
<tr>
<td>Parliament</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 3. The set of concepts that directors communicated to matchers (used in Experiments 1–4).
4.1.2. Participants
Twenty-four undergraduate participants took part in the SD-F condition. Twelve of them acted as directors, and the other 12 matchers subsequently identified the drawings after all had been drawn. Thirty-two undergraduates (working in pairs) took part in the SD+F condition. Another 32 undergraduates (working in pairs) took part in the DD+F condition. Each participant was paid £5.

4.1.3. Data analysis

4.1.3.1. Graphical complexity: To quantify graphical complexity we adopted the Perimetric Complexity (PC) measure first defined by Attneave and Arnoult (1956) and subsequently developed by Pelli, Burns, Farrell, and Moore (2006). (PC = Inside + Outside Perimeter²/Ink Area.) This measure is obtained by first identifying the inside and outside perimeter of the graphic by removing all pixels not on an edge, squaring the result and then dividing it by the ink area (i.e., all the black pixels in the graphic). PC has been identified as a scale-free measure of the graphical information in a picture, and Pelli et al. have convincingly demonstrated that the inverse of the PC of a graphic predicts the efficiency with which observers can identify the graphic from a learned set of competitors. For example, they show that the PC of a letter in a particular font is inversely correlated with the efficiency of identifying the letter in that font when masked with Gaussian noise (log efficiency correlates with log complexity with r = −0.90).

To illustrate the relation between PC and perceived complexity examples of drawings together with their PC are shown in Fig. 4. In general, the PC measure does a good job of capturing intuitive judgments of graphical complexity (see drawings in Fig. 4a). Taking a subset of the drawings (268 drawings taken from the first and last drawings of pairs in the DD+F condition), we had a naive participant rank the members of each pair of drawings in terms of perceived graphical complexity and then compared those rankings with the rankings produced by Pelli et al.’s (2003) PC measure (continuous measure transformed to rank order). There was 82% agreement between the two sets of rankings (kappa = 0.64, p < .001). Examining the 18% cases where there was a discrepancy in the rank ordering suggested that the human observer treated graphics with filled areas as more complex than those with unfilled areas (see Fig. 4c for an example of such a discrepancy due to the sunglasses) and graphics with more distinct meaningful elements as more complex than those with less distinct meaningful elements (see Fig. 4b). In other words, when a picture has two distinct meaningful elements, such as the man and the roulette table in Fig. 4b (left), it is judged more complex than when it only has one meaningful element as when there is only the roulette table, in Fig. 4b (right). Such semantic complexity is not reflected in the pure PC measure. Because we are interested in establishing only the structural complexity of the graphics as opposed to their semantic complexity (see Final Discussion for a detailed discussion of this), we take PC as an objective measure of the graphical complexity of the drawings.

4.1.3.2. Feedback: The drawings from each trial in the SD+F and DD+F conditions were analyzed for evidence of graphical feedback. (It was possible to identify instances of feedback from the different colored inks used by directors and matchers in the task.) The feedback was then classified according to whether it was minimal (a tick) or more interactive (involving one
Fig. 4. Examples of first and last drawings from pairs in the Double Director with Feedback condition of Experiment 1 together with their Perimetric Complexity (PC) scores. Note: In 4a, the PC scores agree with graphical complexity judgments from a human rater. In 4b and 4c, they do not (see text).

4.1.4. Results
4.1.4.1. Overall task performance: For the two feedback conditions (SD+F, DD+F), identification accuracy was based on the performance of the communicators themselves. For the no-feedback SD-F condition, identification accuracy was based on the performance of

or more rounds of annotation from both director and matcher). This data was used to measure the extent of feedback across different conditions.
the second group of 12 matchers who were paired with a particular director and given the sequence of drawings produced by that director. Fig. 5 shows the identification rates (as proportion correct) for the SD-F, SD+F, and DD+F conditions across blocks.

As can be seen from the figure, identification accuracy improved across blocks in all three conditions. However, overall identification rates were lower in the SD-F condition than in the SD+F or DD+F conditions. Finally, unlike the SD+F condition, where identification accuracy plateaus with a slight fall after Block 4, in the DD+F condition matchers maintain identification performance across the final two blocks of the experiment.

Linear trend analysis of variance (ANOVA) on the error proportions confirms these observations. Both by-subject ($F_1$) and by-item ($F_2$) analyses were conducted. $F_1$ trends analyses on Block (1–4 or 1–6) with Condition as a separate factor used a $3 \times 4/6$ mixed design ANOVA, with Block as within subject and Condition (SD-F, SD+F, and DD+F) as between. $F_2$ analyses were more powerful, treating both Block and Condition as within-subjects factors. For all effects reported in this article $p < .05$ unless otherwise stated.

The first set of analyses, conducted over Blocks 1 to 4, produced a reliable linear trend for Block, $F_1(1, 41) = 34.86$ and $F_2(1, 15) = 11.39$; and a main effect of Condition in the by-items analysis, $F_2(1, 15) = 9.71$; but not in the less powerful by-subjects analysis ($p > .1$). There was no interaction between Condition and Block. Further analysis of the Condition effect revealed that there was reliable difference by-items between SD-F and DD+F, $t_2(15) = 3.36$, $p < .01$; and also between SD-F and SD+F, $t_2(15) = 2.57$, $p < .05$; but not between SD+F and DD+F.

An additional trends ANOVA was conducted across the six blocks of the SD+F and DD+F conditions. There was a reliable linear trend for Block, $F_1(1, 30) = 25.87$ and $F_2(1, 15) = 8.59$. There was also a main effect of Condition in the by-items analysis, $F_2(1, 15) = 8.98$. This reflects the greater accuracy of those in the DD+F condition over the SD+F condition. However, there was no interaction between Block and Condition ($Fs < 1$).
4.1.4.2. Graphical complexity: Fig. 6 gives examples of sequences of drawings from the SD-F, SD+F and DD+F conditions. It is clear from these examples that the refinement in drawings across blocks is quite different for the three cases. This difference can be noted in the PC measures shown in Fig. 7. (Mean scores were calculated after removal of scores 2.5 SDs [2% of data] above or below the condition mean and replaced with the relevant cutoff scores.)

As can be seen from the figure, in contrast to the SD-F condition where graphical complexity increases across blocks of the experiment, in the SD+F and DD+F conditions graphical complexity decreases. Furthermore, graphical complexity starts at, and decreases at, a similar rate in these last two conditions.

Linear Trend ANOVAs on the PC measures confirm these observations. Both by-subject ($F_1$) and by-item ($F_2$) analyses were conducted. $F_1$ trends analyses on Block (1–4 or 1–6) with Condition as a separate factor used a $3 \times 4/6$ mixed-design ANOVA, with Block as a within subject and Condition (SD-F, SD+F, and DD+F) as between. $F_2$ analyses were more powerful, treating both Block and Condition as within subject factors.

The first set of analyses, conducted over Blocks 1 to 4, showed a reliable linear trend for Block, $F_1(2, 41) = 9.54$ and $F_2(1, 15) = 5.00$. There was also a main effect of Condition, $F_1(2, 41) = 9.54$ and $F_2(1, 15) = 116.94$. More important, there was a reliable interaction between Block and Condition, $F_1(2, 41) = 22.54$ and $F_2(1, 15) = 116.95$. This was due to the increasing complexity of graphical representations in the SD-F condition, $F_1(3, 123) = 8.28$ and $F_2(3, 45) = 18.38$, compared with their decreasing complexity in the SD+F condition, $F_1(3, 123) = 9.50$ and $F_2(3, 45) = 15.56$; and the DD+F condition, $F_1(3, 123) = 6.82$ and $F_2(3, 45) = 10.91$. There was no difference in the complexity pattern between the SD+F and DD+F conditions ($t$s < 1.65).

An additional trends ANOVA was conducted across the six blocks of the SD+F and DD+F conditions. It showed a reliable linear trend for Block, $F_1(1, 30) = 34.61$ and $F_2(1, 15) = 53.47$. It also showed a main effect of Condition in the by-items analysis, $F_2(1, 15) = 6.13$. This reflects the reduced complexity of drawings in the DD+F condition over the SD+F condition. However, there was no interaction between Block and Condition ($F$s < 1).

4.1.4.3. Feedback: Fig. 8 shows an example of a complex sequence of graphical feedback from a matcher and director in the first block of the DD+F condition. The director is trying to portray Robert De Niro (picture 1 in the sequence). In picture 2, the matcher graphically queries the depiction, asking something along the lines, “Is this a drawing of someone who can be seen on television?” The director then emphasizes who is being depicted by circling the head of the “to be identified” representation and adds a star to indicate that this person is indeed a film/television “star” (picture 3). The matcher then queries this depiction by drawing a question mark and an arrow linking the question mark to the “to be identified picture” (picture 4). This indicates that the depiction is inadequate for identification purposes. Next, by drawing a circle with an arrow and using additional arrows to link this to his or her depiction, the director indicates that his or her depiction is of a male (picture 5). But this is still insufficient and the matcher has to draw a gun with a question mark, asking whether the
Fig. 6. An example of the drawing of “art gallery” over the four blocks of the Single Director with no Feedback (SD-F) condition and the six blocks of the Single Director with Feedback (SD+F) condition and of “De Niro” over the six blocks of the Double Director with Feedback (DD+F) condition from Experiment 1.
Fig. 7. Perimetric Complexity of items drawn in the Single Director with no Feedback (SD-F), Single Director with Feedback (SD+F), and Double Director with Feedback (DD+F) conditions of Experiment 1.

Fig. 8. Examples of the seven rounds of feedback needed to identify a drawing of “De Niro” from the first block of a pair in the Double Director with Feedback condition of Experiment 1.

person depicted is “an action hero” (picture 6). Finally, the director replies by drawing three figures (the competing action figures used in the task) and indicating with arrows that it is the mid-sized action hero. This appears to resolve the drawing ambiguity, as the matcher makes his or her selection.
In fact, this kind of complex repair process proved to be extremely rare. Figure 9 shows the proportion of trials with more than minimal feedback (i.e., involving more than just a tick) across all blocks for the SD+F and the DD+F conditions. For the first block, only 17% of trials in the SD+F and 7% in the DD+F were accompanied with more than minimal feedback. By the third block, the percentages were less than 1% in either condition. So, on most occasions matchers were happy to offer the minimal tick against the initial depiction to indicate recognition.

The proportion of trials with more than minimal feedback were entered into linear trends ANOVAs on Block (Games 1–6) with Condition (SD+F, DD+F) as an additional factor. The by-subjects analysis treated Condition as between subject and Block as within, whereas the by-items analysis treated both Condition and Block as within. The analyses showed a reliable linear trend for Block, $F_1(1, 30) = 28.45$ and $F_2(1, 15) = 21.30$; and a main effect of Condition in the by-items analysis, $F_2(1, 15) = 7.07$; but not by subjects. These main effects were qualified by an interaction between Block and Condition in the by-items analysis, $F_2(1, 15) = 20.40$. The interaction reflected the fact that there was more above-minimal feedback in the first Block for the SD+F condition than the DD+F.

Presumably, matchers in the SD+F condition are aware that they will not have an opportunity to draw the objects subsequently so they more readily engage the drawer when the referent of the drawing is unclear. This stands in contrast to those in the DD+F condition who are afforded the opportunity to indicate understanding by adopting the previously used representation or offer an alternative, potentially more effective representation.

**4.1.5. Discussion**

The results from Experiment 1 allow us to rule out a simplification through repeated production explanation for the iconic to symbolic transformation of the graphical signs in the ‘pictionary’ task. In fact, the complexity results from the SD-F condition indicate that drawings become increasingly complex and retain their iconicity with repeated production in the absence of any feedback from matchers (compare the example drawings in Fig. 5).
This result is reminiscent of findings in referential communication tasks. When speakers have to repeatedly describe the same picture to an imaginary audience, their descriptions often become longer and more complex over trials (Hupet & Chantraine, 1992; Krauss & Weinheimer, 1966). One speculation is that directors in the absence of feedback think of more features of the objects/concepts to be communicated on each occasion and include these additional features in each successive representation.

By contrast, the overall pattern of results is consistent with the grounding explanation whereby feedback enables matchers and directors to ground each representation and commit it to memory. On each subsequent occasion the director can then produce a reduced representation sufficient to elicit the prior use of that representation that is then grounded again through feedback. Finally, the results from the performance measure do point to a difference between the effect of comprehension-based feedback alone and the effect of reciprocal production and comprehension. Participants in the DD+F condition continued to improve and/or maintain their performance across the whole six blocks of the experiment, whereas those in the SD+F condition improved at a slower rate (as shown in the by-items analysis). Also, graphical complexity measures were numerically lower for the DD+F than SD+F drawings (as shown in the by-items analysis). This means that the improved performance on the task was not at the expense of increased complexity in the drawings; if anything, quite the opposite. Experiment 1 demonstrates that communicative feedback is required to support the transition from iconic to symbolic form of graphical signs in the ‘pictionary’ situation, as predicted by the grounding hypothesis. Interestingly, in the vast majority of cases, a simple mark of recognition is sufficient for this to happen. More complex feedback was extremely rare (particularly in the DD+F condition). The results also support the hypothesis that reciprocal production and comprehension enhances the mutual intelligibility of the grounded graphical signs. The second experiment was designed to manipulate the quality of interactive feedback available to participants and establish the extent to which quality of feedback affects the rate of simplification of graphical signs.

4.2. Experiment 2

4.2.1. Design and procedure

Experiment 2 had two conditions corresponding to different levels of interaction for the Double Director arrangement in Experiment 1. In the concurrent feedback condition, graphical interaction took place in real time with partners standing side-by-side at the whiteboard. In the non-concurrent feedback condition, a visual barrier separated the partners during drawing; but after each drawing episode was finished, they were able to interact graphically (as in the SD+F and DD+F conditions in Experiment 1). Whereas the concurrent feedback condition simulates face-to-face conversation in which feedback can accompany the speaker’s production, the non-concurrent condition simulates e-mail communication in that participants can only exchange graphics once they have completed them. In both concurrent and non-concurrent feedback conditions, participants alternated between acting as director and acting as matcher and in all other respects the procedure was the same as that for DD+F condition of Experiment 1.
4.2.2. Participants

The participants were 56 undergraduates, each of whom was randomly assigned to concurrent or non-concurrent feedback conditions. Twelve pairs participated in the concurrent feedback condition and 16 in the non-concurrent condition. Members of each pair had never met each other prior to the experiment, and each participant was paid £5 for taking part.

4.2.3. Results

4.2.3.1. Overall performance: As can be seen from Fig. 10, identification rates (proportion of items correctly identified by matchers) in the concurrent and non-concurrent feedback conditions were comparable. In both cases, identification rates improved across Blocks. Proportion scores were entered into $2 \times 6$ trends ANOVAs carried out by subject ($F_1$) and by item ($F_2$). The by-subject analysis used a mixed design, treating Type of Feedback (concurrent, non-concurrent) as a between-subject factor and Block (1–6) as within, whereas by-item tests treated both as within. Analyses showed a reliable linear trend for Block, $F_1(1, 26) = 26.32$ and $F_2(1, 15) = 32.56$; but no effect of Type of Feedback. However, in the by-items there was a marginal interaction between the two factors, $F_2(1, 15) = 4.33$, $p = .055$. This reflected the slightly steeper increase in accuracy across Blocks for the concurrent feedback condition over the non-concurrent. Thus, participants’ ability to identify the referent of their partner’s drawing was only weakly influenced by the nature of the feedback.

4.2.3.2. Graphical complexity: To illustrate the typical changes in representational form across blocks, we provide an example from concurrent feedback conditions in Fig. 11. The first drawings from each participant take an iconic form, resembling the concepts they depict. However, over the course of the experiment directors strip unnecessary information from the initial representation, leaving only the salient aspects of the image. Note also that as a result of the interaction, participants’ drawings converge; contrast the first drawings by Director 1 and Director 2 of computer monitor (i.e., the first 2 drawings in Fig. 11) with the last drawings by the same participants (i.e., the last 2 drawings in Fig. 11).
Fig. 11. Examples of drawings of “Computer Monitor” over six blocks from the Concurrent Feedback (CF) condition of Experiment 2.

The mean PC scores for the drawings were calculated after replacing scores more than 2.5 SDs from the condition mean (1.92% of data). Fig. 12 shows the result across Blocks in the concurrent and non-concurrent feedback conditions (for comparison measures of complexity for the SD-F condition of Experiment 1 are also shown).

The figure shows that the complexity of drawings decreases more rapidly and more smoothly in the concurrent feedback condition than the non-concurrent feedback condition. These observations are corroborated by trends ANOVAs.

Complexity scores were entered into $2 \times 6$ trends ANOVAs. By-subject analysis used a mixed design, treating Type of Feedback (concurrent, non-concurrent) as a between-subject factor and Block (1–6) as within, whereas by-item analysis treated both factors as within. There

Fig. 12. Perimetric Complexity of items in the Non-Concurrent Feedback (NCF) and Concurrent Feedback (CF) conditions of Experiment 2 (the complexity scores for the Single Director with no Feedback [SD-F] condition of Experiment 1 are shown for comparison).
was a reliable linear trend for Block, $F_1(1, 26) = 45.87$ and $F_2(1, 15) = 65.86$; as well as a main effect of Feedback in the by-items analysis, $F_2(1, 15) = 16.7$; marginally reliable by-subjects, $F_1(1, 26) = 2.85$, $p < .1$. These effects were qualified by a reliable interaction in the by-items analysis, $F_2(1, 15) = 19.02$; marginally reliable also by-subjects, $F_1(1, 26) = 3.05$, $p < .1$. As with the accuracy measures, these effects indicate that the drawings become increasingly simplified and the simplification proceeds faster in the concurrent feedback condition than in the non-concurrent.

4.2.3.3. Feedback: As in Experiment 1 there were very few trials in which there was more than minimal feedback given. Even in the first block only 7% of trials in the non-concurrent, and the concurrent feedback conditions required more than a tick to secure recognition. After this, the level dropped to around 1% of trials. Thus, the increased rate of simplification of the drawings in the concurrent condition as compared to the non-concurrent condition can only be attributed to the improved nature of the feedback (i.e., that it can be given during the drawing process) rather than to the amount or complexity of the feedback given. Because the proportion of more than minimal feedback trials was so low, ANOVA was considered inappropriate for this data.

4.2.3.4. Graphical convergence: We examined graphical convergence in the drawings produced in the concurrent and non-concurrent feedback conditions. For each pair, the drawings for each item (e.g., drama, as drawn by each subject in the pair) were placed in three categories: early (Block 1 & 2 images), middle (Block 3 & 4 images), and late (Block 5 & 6 images). Two independent judges then rated each pair of images (presented in random order) for similarity on a 10-point scale ranging from 0 (dissimilar) to 9 (very similar). The intercoder agreement was high with a correlation of $r = 0.85$ between the similarity judgments for each pair of drawings. The mean similarity judgments for early, middle, and late drawings from the pairs are shown in Fig. 13.

Fig. 13. Similarity ratings from the two judges for early (Rounds 1 & 2), middle (Rounds 3 & 4), and late (Rounds 5 & 6) drawings produced in the Non-Concurrent Feedback (NCF) and Concurrent Feedback (CF) conditions of Experiment 2.
As can be seen from the figure, graphical convergence increased across blocks for both the concurrent and non-concurrent feedback conditions with no apparent difference between the two conditions. These data were entered into linear trends ANOVAs. The by-subjects analysis was a mixed design with Block (early, middle, late) as a within subjects and condition as between subjects. For the by-items analysis, both Block and Condition were treated as within-subject factors. The results confirm our observations yielding a main effect for Block, $F_1(1, 26) = 45.25$ and $F_2(1, 15) = 52.17$. No other effects or interactions approached significance (all $F$s < 1).

Thus, as in verbal communication in which conversational partners entrain one another’s verbal expressions, partners’ graphical expressions converge through interaction (Brennan & Clark 1996; Garrod & Anderson, 1987; Garrod & Doherty, 1994). Interestingly, the degree of convergence was not influenced by the quality of feedback.

4.2.4. Discussion

The results from Experiment 2 extend the findings from Experiment 1 in two ways. First, comparison of the relative reduction in graphical complexity of drawings indicates that the more direct the feedback between participants the greater the tendency to reduce the complexity of the drawings over the course of the experiment. This increased graphical refinement occurs without any loss in identification accuracy because there was no difference in the overall performance of the two groups and a hint that those in the concurrent feedback condition increased accuracy at a greater rate than those in the non-concurrent feedback condition. The second finding is that the drawings of each director in both concurrent and non-concurrent feedback conditions increasingly converge.

These results are interestingly comparable to those found in other referential communication tasks (Clark & Wilkes-Gibbs, 1986; Hupet & Chantraine, 1992; Krauss & Weinheimer, 1966). In the same way that referential descriptions become increasingly shortened in the presence of listener feedback, graphical signs become increasingly simple; and, in line with our earlier argument, they become more symbolic. Similarly, in the absence of any listener feedback, referential descriptions become more complex with repetition (Hupet & Chantraine, 1992) and so do graphical signs. Furthermore, in line with our earlier arguments, they retain their iconicity. We attribute this pattern to a general interactive communication mechanism that is in common to the two situations. As communicators ground the signs successfully, they can increasingly leave out non-essential information. Notice that such collaborative reduction in the sign’s complexity depends upon the prior history of use of that sign by those communicators. In the referential communication literature, the importance of this shared history of interaction was shown by comparing an ‘overhearer’s’ ability to understand the referential description with that of a conversational partner. Schober and Clark (1989) demonstrated that ‘overhearers’ performed systematically less well than conversational partners when it came to identifying the intended referent of a description.

The third experiment was designed to establish whether ‘overseers’ would have the same problem with graphical signs produced in the ‘pictionary’ experiments as ‘overhearers’ do with referential descriptions.
4.3. Experiment 3

In this experiment ‘overseers’ were given complete sets of drawings taken from those produced by the participants in the SD+F and DD+F conditions of Experiment 1. The drawings were presented in the order in which they had been produced during the original experiment, and ‘overseers’ were required to identify each drawing in relation to the list of alternatives presented to the original Matchers in Experiment 1. There were two groups of ‘overseers’ that corresponded to Schober and Clark’s (1989) early and late ‘overhearers.’ The early ‘overseers’ saw all the drawings produced starting with Block 1 and moving through to Block 6, whereas the late overseers were only exposed to drawings produced in the final three blocks (Blocks 4–6). The prediction is that both early and late ‘overseers’ will perform consistently less well than the original matchers, because they cannot engage in any interaction with the original director to help clarify the meaning of the drawing. Furthermore, late ‘overseers’ should perform consistently less well than early ‘overseers.’ This is because they do not have access to the full communicative history of that drawing. As a consequence, they will be in a less favorable position to establish its symbolic interpretation than the early ‘overseers.’

4.3.1. Method

Twenty-eight complete sets of drawings were sampled from 14 pairs of players in the SD+F and 14 pairs of players in DD+F conditions of Experiment 1. Each set formed a stimulus set for a different participant. Twenty-eight participants took part in the early overseer with 14 in the SD+F condition and 14 in the DD+F condition. Each participant identified a set of images generated across all blocks (1–6) from one pair of either SD+F or DD+F drawers. In the late overseer condition, 28 different participants did the same for drawings from Blocks 4 to 6. Again each participant identified only one set of images within the conditions. In both conditions, participants were given a table of items, including four distracter items, and were instructed to choose which object they thought the picture represented.

4.3.2. Results

In addition to the original identification rates (taken from the SD+F/DD+F conditions of Experiment 1), early ‘overseer’ (SDearly for the SD+F ‘overseers’ and DDeary for the DD+F ‘overseers’) and late ‘overseer’ (SDlate and DDLate) identification rates are plotted in Fig. 14. As hypothesized, ‘overseers’ performed consistently less well than the original matchers across all blocks. Identification accuracy was better in the DD+F condition than for both early/late ‘overseers’ (DDeary, DDLate), and early ‘overseers’ were better at identifying the items than late ‘overseers.’ Similarly, identification accuracy was better for the SD+F condition than for both early and late ‘overseers’ (SDearly, SDlate), and early ‘overseers’ were better than late ‘overseers.’

Linear trends ANOVAs were carried out to confirm the observed differences. Because each participant had a different set of the original stimuli (i.e., all items in the population of drawings sampled were assigned to different subjects on a one-to-one basis) only by-subjects ANOVAs are appropriate (Clark, 1973). The first trends ANOVA was on Block (1–6; within)
Fig. 14. Identification accuracy for original matchers in the Single Director (SD) with Feedback and Double Director (DD) with Feedback conditions of Experiment 1 compared with that of early overseers (SDearly/DDearly) and late overseers (SDlate/DDlate) from Experiment 3.

The analysis showed significant main effects of Block, $F(1, 56) = 52.4$; Condition, $F(1, 56) = 3.97$; and, more important, Matcher, $F(1, 56) = 74.44$; but no interactions ($F < 1$).

The second linear trends ANOVA analyzed Block (4–6) with Matcher (Original, Early Overseer, Late Overseer) and Conditions (SD, DD) as additional factors. It showed significant main effects for Condition, $F(1, 82) = 4.55$; and Matcher, $F(1, 82) = 50.37$; with no main effect of Block. Planned comparisons show that the accuracy of original matchers was significantly greater than for early overseers, $t(82) = 6.56, p < .001$; and the accuracy for early overseers was significantly greater than for late overseers, $t(82) = 3.12, p < .01$.

4.3.3. Discussion

As hypothesized, both early and late overseer groups showed similar patterns of identification accuracy to that of the original matchers only at significantly lower accuracy rates. This is in line with Schober and Clark’s (1989) finding that ‘overhearers’ who did not participate in the interaction were significantly less accurate in identifying referents than original interlocutors. In addition, late ‘overseers,’ not exposed to the original interaction and refinement of the drawings, found identifying later symbolic drawings significantly more difficult than both original matchers and early overseers.

The results from the overseer experiment support the hypothesis that the drawings are refined as a result of feedback given in the original SD+F and DD+F situations. Without being able to engage in the interaction itself, overseers are much poorer at interpreting the drawings than the original participants. This is exactly as predicted by our hypothesis that the locus of structure moves from the sign into the communicators’ memory representations as a consequence of grounding the signs during the course of the interaction. In this respect, performance in the graphical communication task is strikingly similar to that observed in referential communication tasks using speech. This indicates that the simplification and convergence of
pictorial representations is driven by the same interactive constraints that apply in interactive language use.

5. Final Discussion

This article set out to investigate the communicative origin of graphical symbols. It is straightforward to explain how iconic graphical representations arise because their interpretation is based on perceptual resemblance to their objects. It is not so obvious how graphical symbols arise because their interpretation depends upon prior understanding of the arbitrary relations between the sign and its object. However, if it can be shown that graphical symbols systematically evolve from the repeated use of iconic and indexical signs, then this would account for the problem. Although such an evolution is suggested by the historical development of many symbolic writing systems, which start out as iconic representations, it is not clear precisely what drives the evolution.

The results from the experiments clearly show that in the case of the ‘pictionary’ task this evolution depends upon interaction between communicators. Experiment 1 showed that in the SD-F condition in which there was no interaction between director and matcher, the drawings became increasingly complex with repetition and retained their iconic character, whereas with even minimal opportunity for feedback in the SD+F condition the drawings became increasingly simple and more symbolic in character. Furthermore, the experiment showed that when both communicators are able to produce the drawings (DD+F condition) the communication becomes more effective. Matchers become better at identifying the drawings, whereas the drawings become at least as simple if not slightly simpler than in the SD+F condition. Experiment 2 showed that the rate at which the drawings become simpler and more abstract is affected by the quality of the interactive feedback allowed by the task. This experiment also demonstrated that the drawings produced by each communicator increasingly converge with those of their partner. Experiment 3 demonstrated that interpreting the drawings appropriately depended on being party to the interaction that drives their simplification and abstraction. Overseers not engaged in the original communication process perform consistently less well at identifying the drawings than the original participants. According to our argument in section 1, this process of simplification and abstraction reflects a transformation from iconic graphical signs to symbolic graphical signs.

Of course, we should make it clear that these experiments do not rule out the possibility that symbolic graphical representations may also emerge out of extended non-interactive and repeated use. However, what they do show is that the transition from iconic to symbolic graphical signs occurs very rapidly in an interactive communication setting and given the opposite pattern of results for the non-interactive communication setting, it seems unlikely that the move from iconic to symbolic graphical representations could evolve rapidly without this kind of interaction. We can speculate that there may be an alternative grounding mechanism that could operate with repeated and reciprocal use of graphical signs in a community even when there is no direct interaction (e.g., as in the use of a writing system). This is because recurrent and reciprocal usage of signs offers some feedback to ground the signs (e.g., I have produced a sign that seems to work because it is the same as that being used by others).
However, crucially, this situation still requires evidence of grounding through indirect feedback from other sign users.

So, in summary, our graphical communication studies offer an insight into the evolution of graphical symbols from iconic and indexical signs. Also they are consistent with the information theoretic account of the basic nature of signs. In this account, iconic (and indexical) signs are inherently more complex than symbolic signs because they need to reflect the structure of their objects in the structure of the sign itself according to the structure mapping principle. Symbolic signs only require enough structure to enable users to identify that sign from its competitors in the set. However, this brings us to an aspect of signification apparent in many of the drawings produced in the ‘pictionary’ experiments that we have not considered thus far, which introduces the problem of compositionality and systematicity of signs.

5.1. Compositionality and systematicity of signs

Consider the series of drawings for “art gallery” shown in Fig. 15, which were produced by a pair in Experiment 1. Like many of the graphical depictions encountered in the ‘pictionary’ experiments, all of the examples are semantically complex. In other words, they contain distinct graphical elements each with its own semantic interpretation, which in turn combine with the interpretations of the other elements to produce an interpretation of the whole. Each drawing contains three different elements: (a) an iconic depiction of a building, (b) a pair of diverging lines similar to a “less than” symbol, and (c) an iconic depiction of a painting. The interpretation of the whole drawing is derived from interpreting the less than symbol as indicating that the object signified on its left contains the object(s) signified on its right: Hence, it is a building that contains paintings. With repeated use, the drawings become graphically

![Fig. 15. Depictions of art gallery from a pair in Experiment 1 illustrating increasing simplicity without reduction in semantic complexity.](image-url)
simpler, and the iconic components depicting the building and the painting become more symbolic; nevertheless, the whole sign retains its semantic complexity.

Now, compare two hypothetical sets of signs: one containing semantic complexity like this, and the other without it. As argued in the introduction, we can move structure from individual signs into representations that reflect the history of interaction. We can also put it into the structure of the coding system itself. This means that we can detach the graphical complexity of a sign (its lines and details) from its semantic complexity. The latter corresponds to the extent to which the sign contains independent subcomponents. When these subcomponents recur across multiple signs, each sign associated with a different object, there is structure within the coding scheme. When a subcomponent only occurs in signs associated with members of a family of related objects, then the coding scheme’s structure corresponds to (indeed, structure-maps) at least some of the structure within the domain of objects. For example, for the pair producing the art gallery signs in Fig. 15, the building subcomponent became associated with all and only the institutions in the domain of objects (i.e., museum, art gallery, theatre). This is very different from the structure mapping found in icons: There, subparts of iconic signs correspond to subparts of the objects they resemble. Here, subparts of signs need not correspond to subparts of objects. Rather, subparts of the signs correspond to taxonomic subparts of the domain of objects as a whole (e.g., the subpart that corresponds to all and only buildings in the domain of objects in the pictionary task).

Thus, a system of signs can contain complex signs whose meaning depends upon subcomponents that also occur in other complex signs. We refer to this property as systematicity. It seems plausible that naturally occurring sets of signs can mix systematic with non-systematic signs: Systematic signs can contain both iconic and symbolic subcomponents.

Our account of the evolution of sets of icons into sets of symbols, and of sets of symbols into symbol systems has some direct parallels in Deacon’s (1997) Peircean account of signs. However, there are some terminological differences, and Hurford (1998) argues forcefully that Deacon’s terminology is non-standard. For example, where we would use the term symbol, Deacon reserves the term index for a sign regularly associated with an object. Whereas we consider a set of such associations to constitute a symbol system, Deacon additionally requires the properties we have associated with systematicity (p. 99). But, in spite of these differences, there are several parallels. For instance, Deacon too emphasizes how perceptual categorization, memory, and prediction must interact to lead to fully developed symbol systems (p. 78); and he emphasizes that the set of relations between signs can come to resemble the set of relations between objects (p. 88).

There remain some deeper differences. For example, when Deacon (1997) discusses symbolic systems, he assumed that an object is (indexically) picked out by a purely arbitrary sign. The assumption of arbitrariness drives a distinct wedge between sets of icons and sets of symbols. Here, we have emphasized that sets of symbols may evolve from sets of icons; and hence, members of a set of symbols may possess a significant degree of residual iconicity. It means that whether a set of signs is systematic and whether its signs retain iconicity, are orthogonal questions.

The idea that systematicity and iconicity vary independently fits with a number of other theories and observations. For instance, Barsalou’s (1999) perceptual symbol systems theory makes room for mental representations composed of parts that perceptually resemble objects.
And, in natural language, there has been continuing interest in the extent to which supposedly arbitrary compositional systems possess iconic features. For instance, structure in syntax and discourse often iconically represents temporal structure or task structure (Haiman, 1985). More recently, Shillcock, Kirby, McDonald, and Brew (2001) argued that resemblances between words in the English lexicon are non-arbitrary, so that perceptual similarities in phonological form can be a guide to underlying similarities in meaning, and vice versa. And, work on American Sign Language (ASL) has argued that such non-arbitrary “richly grounded” forms can co-exist with arbitrary forms (Macken, Perry, & Haas, 1993). Furthermore, ASL can be compared with home-sign languages, where deaf individuals in non-signing households develop their own communication systems (e.g., Goldin-Meadow & Mylander 1998). Reviewing such work, Morford (1996, pp. 166–167) suggested that iconicity is retained when “symbols are used with individuals unfamiliar with the system.”

Thus, we advocate further investigation of the ways in which icons can evolve into symbols, the constraints on simplification-by-grounding, and the ways in which symbol systems can sustain residual iconicity. Previous work has studied naturally occurring systems of signs; on the basis of our experience, we would argue that the ‘pictionary’ paradigm we have exploited here provides an excellent vehicle for further, controlled investigation of the evolution of representations.

Notes

1. Tversky (1995) included additional factors in her production-driven account such as the changes in writing media (e.g., from stone, to papyrus, to hand-held clay tablets) that also attended the changes in the writing systems that she wrote about.

Acknowledgments

This research was supported, in part, by the National Institute of Information and Communications Technology, Japan; the ESRC (Grant L323253003); and the ARC (Grant DP0556991).

References