slipping the default height of 'l' is thrown out in favor of slipping the dot of 'i' out of existence or perceiving the serif of the 'l' as the dot of an 'i'.

Prediction 4

Figure 10 shows a plot of the number of categories in which each token of **ALL** was found (including the correct category) vs. its average reaction time. There is a strong correlation here (r=0.85, p<0.01). Low RT's occur when there are few possible category bins. As the number of possible answers increases, RT becomes longer.

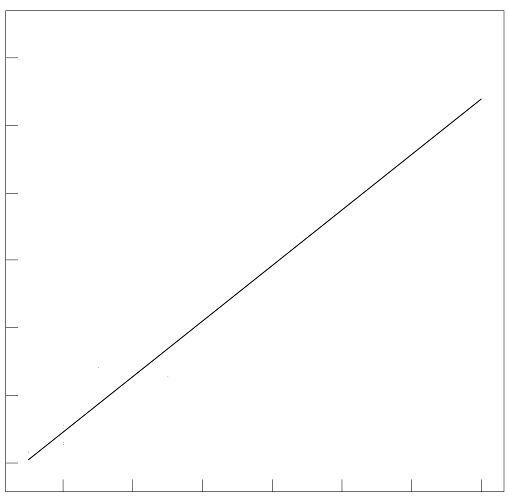

Figure 10: Number of categories in which each token of **ALL** was found (including the correct answer) vs. reaction time.

This result can be used to classify the tokens into two sets: those with *ambiguous* role fillers and those with *strange* role fillers. Ambiguous tokens are those that have a small number of possible categories but more than one. They are found on the left of Figure 10. On the other hand, strange tokens have many possible categories and are found on the right side of the graph. There is a continuum between ambiguity and strangeness.

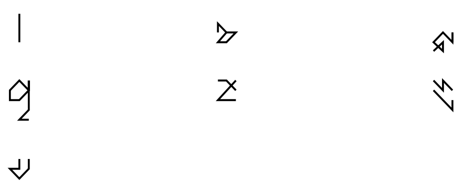

Figure 11: Ambiguous and strange letterforms (14-gous being the strangest of all). Beside each gridletter are the categories people guessed and the average reaction time.

When a token is ambiguous, its parts can fill two or more sets of roles all in a fairly reasonable fashion. This is illustrated by the first two columns of data in Figure 11. A choice must be made between the possibilities. This choice seems to take longer as more competing categories are considered. When a token is strange, its parts do not really fit any sets of roles well (although clearly they will fit some letter-conceptualizations better than others). No answers come to mind even after an extended period or time, and finally a guess of some sort is made. The two tokens in the "14-guous" column of Figure 11 are very strange tokens indeed, each eliciting 14 possible categorizations.

Conclusion

Results from our experiment in human letter recognition provide evidence for the existence of conceptual-level representations of letter parts which we call roles. We intend to incorporate such representations into a computer model of letter recognition. A more complete report on this research is available in (McGraw et al., 1994).

References

Blesser, B., Shillman, R., Cox, C., Kuklinski, T., Ventura, J., & Eden, M. (1973). Character recognition based on phenomenological attributes. *Visible Language*, 7(3).

Gilmore, G., Hersh, H., Caramazza, A., & Griffin, J. (1979). Multidimensional letter similarity derived from recognition errors. *Perception & Psychophysics*, 25(5), 425–431.

Goldstone, R., Medin, D., & Gentner, D. (1991). Realtions, attributes, and the non-independence of features in similarity judgements. *Cognitive Psychology*, 23(2), 222–264.

Hofstadter, D. (1979). *Gödel, Escher, Bach: an Eternal Golden Braid.* New York: Basic Books.

Hofstadter, D. & McGraw, G. (1994). *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*, chapter 9: Letter Spirit: Esthetic Perception and Creative Play in the Rich Microcosm of the Alphabet. Basic Books: New York. To appear.

Kahneman, D. & Miller, D. (1986). Norm theory: comparing reality to its alternatives. *Psychological Review*, 93, 136–153.

Mantas, J. (1986). An overview of character recognition methodologies. *Pattern Recognition*, 19(6), 425–430.

McGraw, G., Rehling, J. & Goldstone, R. (1994). Roles in Letter Perception: Human data and computer models. Technical Report 90, Center for Research on Concepts and Cognition, 510 N. Fess, Bloomington, IN, 47408.

Palmer, S. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology*, 9, 441–474.

Palmer, S. (1978). Structural aspects of visual similarity. *Memory & Cognition*, 6(2), 91–97.

Treisman, A. & Paterson, R. (1984). Emergent features, attention, and object perception. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 12–31.

extent h↔k and v↔w. Note, however, that all of these letter pairs share many common roles as well — especially the first three: 'g' and 'q' both share a bowl or a circle in the central zone and have a descender on the right, 'i' and 'j' differ only in that 'j' has a descender, 'f' and 't' both share the crossbar role and have closely related stems. In fact, we suggest that it is the role similarities that give rise to the shape proximities in these cases, and that confusion in these cases is a result of higher-level role similarities.

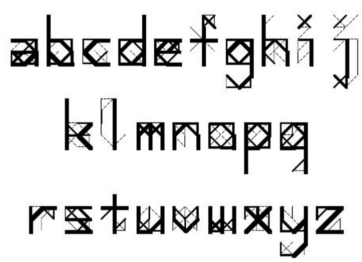

Figure 8: The blurred prototypes of **ALL**. Thicker marks indicate the prevalence of a given quantum.

Error	Percent
u↔v	16.30
g↔q	10.95
i↔l	7.95
i↔j	7.45
f↔t	5.75
v↔w	4.75
h↔k	4.50
n↔r	4.05
c↔r	4.05

Table 3: The top nine bi-directional category errors made by humans.

The cases where cluster proximity differs from human errors are more interesting. First, consider the clusters in Figure 7 that do not correctly predict errors in human perception. The following clusters have very close proximity in blurred-prototype space but only rarely occur as errors (shown in parenthesis): e↔s (0.20%), c↔o (0.65%), u↔w (0.85%), b↔h (1.30%), x↔z (2.35%) and m↔n (2.70%). Clearly, people do not make these errors even though the letters are similar at the lowest of levels.

Our results corroborate the work of (Goldstone et al., 1991) who show that literal physical properties are often less important in similarity judgements than are higher-level relations. This result is compatible with all accounts of categorization that consider emergent features (*e.g.* closure) (Treisman and Paterson, 1984), including the role hypothesis discussed here.

Prediction 3

We hypothesize that there are salient differences in the role sets of these letters that prevent people from confusing them. The difference between 'b' and 'h' is a matter of closure at the bottom. This difference is very salient to people since they know of the existence (and the conceptual proximity) of the competing category. The geometric similarities of the role sets force people to pay special attention to such closure when they are categorizing 'b's and 'h's. Similar stories can be told about 'c' and 'o', 'm' and 'n', and 'x' and 'z'. These letters are rarely confused by people because higher-order *conceptual* differences are too great.

Next, consider a few of the most common errors made by people that are not represented in the clustering. The errors include: u↔v, i↔l, n↔r, and c↔r. Consider the roles for 'r' and 'n'. Both have a stem on the left that goes from the baseline to the x-height. Both also have a role that attaches to the stem near its top on the left and arches over to the right. It is this role that can cause an ambiguity. Depending on how close it gets to the baseline, the amount of 'n'-ness of the letterform will vary. Similar stories can be told for the remaining three common errors not found in the clustering. Because some aspects of roles are slippable, human letter perception is very flexible. This slippability comes at the price of occasional error-making.

Note that *all* of the most common human category errors shown in Table 3 can be understood in terms of role slippages caused by poor role-fillers. First-order syntactic proximity does not provide as powerful an explanation of the results. That errors are easily explained by the role hypothesis becomes even more apparent when specific errors are considered.

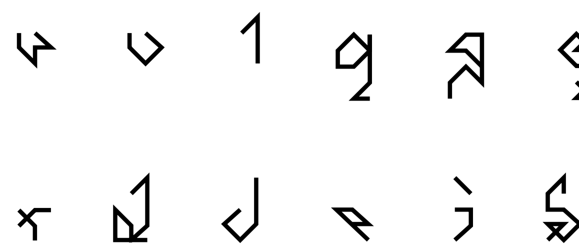

Figure 9: Exampless of some commonly "miscategorized" tokens. Note how reasonable these errors are.

Figure 9 shows a number of commonly misperceived tokens from **ALL**. The first thing to note is that all of the "errors" are fairly reasonable. In fact, it is not clear that some of them are really errors. In any case, most errors can be easily explained in terms of conceptual proximity. All of these errors involve perceptual slippages where unclear role fillers cause miscategorization. In the example shown, the roles of a 'w' are filled quite nicely by what was supposed to be a 'v' — with the kink on the right side being seen as forming the bottom hump of a 'w'. In some cases the perceiver allows roles to slip too much in order to make sense of weird shapes, resulting in miscategorization. This usually occurs with complicated or otherwise non-standard letterforms and is probably magnified under time pressure. In the case of 'l' seen as 'i',

Figure 6: Standard vs. non-standard role-fillers as exhibited in *double backslash* (left letterform of each triplet), *sabretooth* (middle), and tokens from **NORMALS** (right). Starting at the left of each triplet where roles are poorly filled, roles are filled in a more standard fashion towards the right. Role-related difficulty is reflected in the percent correct and RT data.

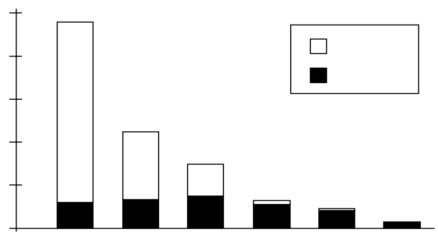

Figure 4: The number of tokens, separated into **NORMALS** and **FONTS** groups, that occurred within certain RT ranges.

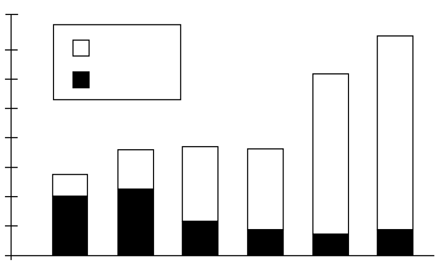

Figure 5: The number of tokens, separated into **NORMALS** and **FONTS** groups, that occurred within certain accuracy ranges.

late highly with the number of quanta (r=0.59, p=0.16), nor does accuracy (r=0.55, p=0.19).[5] That difficulty is more than just quanta number is apparent from the data shown in Table 2 (note especially *close* and *hint four*). The role hypothesis can account for this result as follows: some gridletters with a higher than average quanta count still have well-filled roles while others having an average number of quanta, or possibly even a lower than average quanta count, have troublesome role-fillers. This effect is also apparent within the **FONTS** group, with *sabretooth* having fairly strong but heavy role-fillers (making it relatively easy to recognize) and *double backslash* having roles filled in less standard ways (making it harder to recognize). Most **NORMALS** tokens, having very normal role fillers, are even easier to recognize than the *sabretooth* tokens. Figure 6 shows some specific tokens (as well as accuracy and RT data) to illustrate this point.

[5] Again, correlations are for all tokens averaged over all subjects.

Prediction 2

Figure 7 displays the tree resulting from a hierarchical cluster analysis of the dataset **ALL**. The clustering is meant to provide some insight into the first-order syntactic similarities between letter categories. It is based on quanta-level information alone and is thus very low-level. The clustering is made using the euclidean distances between the *blurred prototypes* of **ALL**. These prototypes, graphically represented in Figure 8, were created by averaging the binary quanta lists of all of the letters in a given category together, resulting in a 56 dimensional vector for each category.

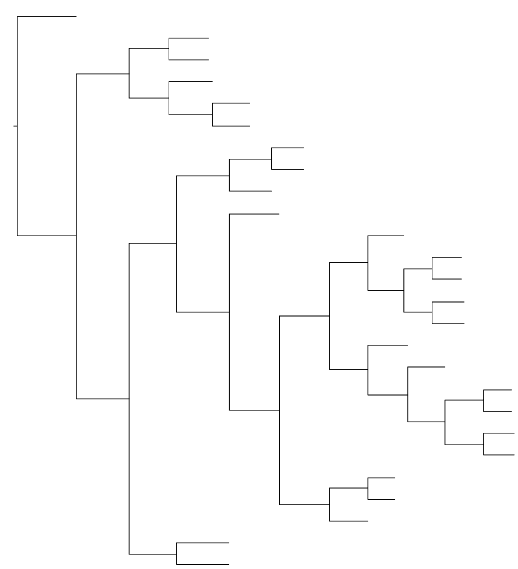

Figure 7: A hierarchical cluster analysis of the 56-dimensional blurred prototypes of **ALL** providing a low-level syntactic view of the dataset.

It is interesting to contrast the clustering shown in Figure 7 to the human category errors in Table 3. Although some human errors seem to be related to euclidean proximity between letter shapes, many are not. The human errors that can be explained or predicted in terms of euclidean proximity (considering only pairs in the clustering as truly close) are g$\leftrightarrow$q, i$\leftrightarrow$j, f$\leftrightarrow$t, and to a lesser

	response																									
stim	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
a	0.833	0.005	0.000	0.026	0.019	0.000	0.011	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.018	0.000	0.011	0.003	0.002	0.002	0.010	0.011	0.018	0.011	0.000	0.013
b	0.002	0.910	0.000	0.005	0.002	0.000	0.002	0.005	0.000	0.002	0.005	0.009	0.000	0.000	0.000	0.003	0.000	0.000	0.046	0.000	0.002	0.005	0.000	0.002	0.000	0.002
c	0.014	0.000	0.884	0.002	0.012	0.000	0.025	0.000	0.002	0.000	0.000	0.006	0.000	0.000	0.010	0.000	0.000	0.002	0.002	0.000	0.012	0.017	0.008	0.004	0.000	0.002
d	0.035	0.011	0.000	0.884	0.000	0.000	0.013	0.000	0.000	0.030	0.002	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.008	0.000	0.003	0.003	0.002	0.002	0.000	0.003
e	0.005	0.000	0.012	0.001	0.912	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.014	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.010
f	0.002	0.000	0.000	0.002	0.034	0.732	0.005	0.025	0.009	0.018	0.002	0.004	0.022	0.000	0.000	0.007	0.009	0.005	0.007	0.108	0.000	0.002	0.004	0.000	0.002	0.000
g	0.003	0.000	0.000	0.000	0.000	0.003	0.906	0.003	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.003	0.040	0.000	0.025	0.000	0.000	0.000	0.000	0.000	0.004	0.000
h	0.000	0.021	0.004	0.000	0.006	0.002	0.008	0.896	0.000	0.000	0.008	0.000	0.002	0.023	0.000	0.000	0.000	0.000	0.010	0.002	0.002	0.002	0.000	0.002	0.010	0.000
i	0.000	0.004	0.004	0.000	0.003	0.019	0.013	0.000	0.745	0.102	0.001	0.001	0.003	0.001	0.006	0.004	0.004	0.006	0.046	0.017	0.001	0.003	0.004	0.004	0.003	0.001
j	0.001	0.004	0.000	0.001	0.001	0.001	0.008	0.001	0.047	0.862	0.001	0.002	0.000	0.000	0.000	0.000	0.001	0.000	0.061	0.002	0.001	0.001	0.000	0.001	0.003	0.000
k	0.000	0.011	0.005	0.000	0.005	0.029	0.006	0.082	0.000	0.000	0.786	0.006	0.003	0.003	0.000	0.000	0.000	0.002	0.008	0.003	0.003	0.006	0.002	0.037	0.002	0.000
l	0.000	0.029	0.041	0.000	0.002	0.015	0.004	0.004	0.158	0.021	0.004	0.593	0.000	0.000	0.000	0.000	0.004	0.008	0.031	0.033	0.025	0.004	0.000	0.002	0.000	0.023
m	0.000	0.000	0.000	0.000	0.008	0.004	0.000	0.001	0.000	0.001	0.001	0.001	0.928	0.012	0.000	0.001	0.000	0.000	0.001	0.001	0.001	0.000	0.001	0.019	0.018	0.000
n	0.029	0.004	0.002	0.005	0.004	0.002	0.004	0.022	0.004	0.004	0.000	0.000	0.042	0.817	0.005	0.000	0.000	0.011	0.002	0.013	0.000	0.000	0.002	0.016	0.005	0.007
o	0.027	0.015	0.003	0.053	0.008	0.000	0.000	0.000	0.003	0.005	0.002	0.000	0.000	0.005	0.828	0.002	0.012	0.003	0.003	0.000	0.011	0.020	0.000	0.002	0.000	0.000
p	0.000	0.000	0.000	0.000	0.013	0.000	0.017	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.001	0.936	0.020	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.002
q	0.003	0.002	0.001	0.001	0.001	0.000	0.179	0.000	0.001	0.002	0.000	0.000	0.000	0.002	0.002	0.013	0.775	0.000	0.006	0.000	0.000	0.000	0.001	0.000	0.001	0.002
r	0.012	0.005	0.064	0.000	0.011	0.003	0.002	0.006	0.009	0.003	0.003	0.000	0.000	0.070	0.005	0.020	0.002	0.701	0.009	0.012	0.000	0.005	0.000	0.030	0.026	0.005
s	0.004	0.000	0.000	0.000	0.003	0.000	0.004	0.000	0.001	0.001	0.000	0.001	0.000	0.005	0.003	0.000	0.001	0.004	0.935	0.000	0.004	0.010	0.001	0.004	0.005	0.011
t	0.003	0.010	0.024	0.000	0.005	0.007	0.010	0.000	0.007	0.002	0.007	0.007	0.000	0.002	0.002	0.000	0.003	0.012	0.000	0.866	0.003	0.000	0.002	0.015	0.010	0.003
u	0.009	0.003	0.004	0.001	0.003	0.001	0.003	0.001	0.000	0.003	0.006	0.001	0.000	0.001	0.003	0.000	0.000	0.004	0.001	0.003	0.792	0.127	0.014	0.003	0.014	0.000
v	0.000	0.009	0.017	0.001	0.019	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.000	0.003	0.001	0.199	0.654	0.084	0.001	0.000	0.000
w	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.010	0.001	0.001	0.000	0.000	0.000	0.002	0.000	0.001	0.000	0.003	0.011	0.937	0.021	0.002	0.000
x	0.002	0.000	0.000	0.000	0.000	0.003	0.000	0.002	0.000	0.002	0.006	0.000	0.010	0.002	0.000	0.000	0.000	0.003	0.008	0.021	0.000	0.000	0.000	0.929	0.003	0.008
y	0.001	0.000	0.000	0.000	0.003	0.000	0.004	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.004	0.001	0.007	0.001	0.004	0.001	0.965	0.003
z	0.004	0.000	0.023	0.004	0.023	0.000	0.008	0.002	0.004	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.008	0.006	0.014	0.017	0.010	0.008	0.004	0.039	0.031	0.795

Table 1: Confusion matrix of human performance on **ALL**.

over that of speeded response, although subjects were aware that their response times were being recorded.

For 21 of the subjects, the data set was **NORMALS** plus four of the gridfonts from **FONTS** (104 more tokens, for a total of 493). Two additional gridfonts were added for the final 14 subjects only (for a total of 545 tokens). The order of presentation was uniquely randomized for each subject.

Subjects. 35 subjects were drawn from the Indiana University community and included undergraduates, graduate students, and staff. None were involved in work related to fonts of any kind. All subjects were compensated for their participation and were debriefed after the experiment.

Results

Our experiment invloved the presentation of 17,983 trials divided over the 35 subjects. Subjects varied between 63.7% and 95.6% correct responses with a mean of 84.0% correct.

Table 1 shows a Stimulus by Response confusion matrix generated for the data set **ALL** by averaging over the 35 subjects and collapsing tokens into their respective categories. This confusion matrix differs considerably from others like it in the psychology literature in two ways. First, the letters reported on are less typical examples of their categories than are usually used. Second, the matrix includes data about a variety of typefaces instead of just one, some of which have extremely nonstandard letters. This leads to a more interesting set of errors than has been evident in past work. Our working assumption is that the confusability between two letters increases as a measure of their representational similarity (Gilmore et al., 1979).

Prediction 1

Table 2 shows average accuracy and reaction time values for some of the data subsets. For all groups, with the exception of *sabretooth* and *close*, accuracy and speed have corresponding ranks. Generally, as RT gets longer, accuracy decreases (r=0.74, p<0.01 for a correlation over all tokens averaged over all subjects). Both of these measures can be understood in terms of an intuitive notion of *difficulty*, with RT rising and accuracy falling as difficulty rises.

Stimulus	Accuracy	Avg. RT	quanta
ALL	84.0 %	1.84 s	6.80
NORMALS	90.1	1.46	6.97
FONTS	68.7	2.80	6.36
bowtie	52.1	3.79	7.65
double backslash	66.2	2.88	6.00
flournoy ranch	57.4	3.44	7.54
hint four	85.3	1.95	4.92
close	74.8	2.12	5.31
sabretooth	76.7	2.61	6.73

Table 2: Human recognition performance averaged over 35 subjects.

Difficulty is directly determined by the ease with which a letterform can be broken into parts that fill roles. Complicated letterforms take longer to parse into roles, causing RT to rise. Likewise, letterforms that fill roles poorly or ambiguously cause more errors than those that fill roles easily. Figure 2 provides an example of "good" role fillers (the top examples) vs. "bad" role fillers (the bottom examples). The role hypothesis thus explains the relationship between RT and accuracy.

There is a striking difference in RT (and accuracy) between **FONTS** and **NORMALS**. Figure 4 shows a histogram of the number of tokens in the data set that fell into certain RT bins. It shows that most of the tokens in **NORMALS** are recognized quickly, while most of the **FONT** tokens take longer. Figure 5 shows a similar histogram for accuracy. Clearly, most **NORMALS** tokens are recognized more accurately than the **FONT** tokens. The **NORMALS** generally have parts that easily fill roles, whereas **FONTS** have non-standard role-fillers.

The number of quanta in a gridletter does not alone determine its difficulty. As a result, RT does not corre-

Figure 2 shows how letterforms are comprised of parts which correspond to a letter concept's roles. Such parts are perceived under top-down pressure from roles and are sensitive to context. As stated in (Palmer, 1978) (page 96), "components enter into relationships with other components, resulting in larger structural units whose importance supercedes that of its constituents." We hold that most of the "importance" attributed to the emerging parts stems directly from their role-filling ability. In other words, the way in which a part fills a role directly determines its "goodness."

Some aspects of roles can be *slipped* to accomodate a letterform's parts.[3] "Slippage" involves allowing certain descriptions in a mental representation to "slip," or be replaced by related descriptions according to contextual pressures brought to bear by the situation. Some of the descriptions making up the representation of a role are more slippable than others. The fluid nature of human letter perception is a consequence of slippability at the role-level.

The Gridfont Domain

Our aim in this paper is to show that roles play an important part in letter perception. This becomes especially apparent when a letter is complicated or near the boundaries of its intended category. We are motivated by the need for a reasonable cognitive model of letter perception which we will use as part of the Letter Spirit model of creativity in letter design.[4]

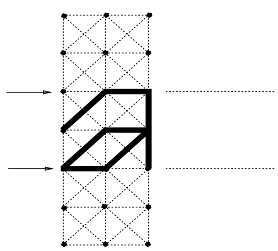

Figure 3: The Letter Spirit grid, with one of the possible sets of quanta instantiating an 'a' turned on.

In the Letter Spirit domain, letterforms are restricted to short line segments on a fixed 3×7 grid. There are 56 possible legal line segments, called *quanta*, that connect a point to any of its nearest neighbors (see Figure 3). Because quanta are either on or off, decisions on the grid are coarse. Surprisingly, the variety among letters of a given category is still huge — hundreds of versions of each letter and around 600 full gridfonts have been designed by humans. Almost paradoxically, the domain's limitations engender this diversity.

The low-level constraints defined by the grid provide a rich microdomain that focuses attention on cognitive issues of letter perception. One important feature of the grid is that it can both force letters to their categorical extremes because of its sparseness and yet still contain many easily identifiable letterforms. This makes it an excellent vehicle for letter perception research — one that allows some attention to be focused on issues of style.

[3]The term "slippability", coined by Hofstadter (1979), is roughly equivalent to the more pervasive psychological term "mutability" used by (Kahneman and Miller, 1986).

[4]See (Hofstadter and McGraw, 1994) for a description of the Letter Spirit project.

The Human Study

The purpose of our experiment is twofold: to gather data for comparison with cognitive models of letter perception, and to shed light on the human perception of lowercase roman letters that are non-standard in nature. Our central predictions all relate to the hypothesis that letter concepts are comprised of simpler conceptual *roles*. Because of the existence of roles, human letter perception should be affected in the following ways:

1. Reaction time should be longer and recognition rate lower for complicated stimuli in which roles are either ambiguously filled or poorly filled than it should be for relatively normal letters with standard role-fillers.
2. First-order syntactic proximity (*i.e.,* proximity of letter shapes considered only at the level of quanta) alone should not account for errors in human letter recognition. Instead, errors should be explainable in terms of role slippages, part-role fillers, and role-set relations.
3. Category errors made by human subjects should be most prevalent between letters with similar sets of roles — roles which are related to each other spatially in a similar fashion. Thus, 'n's and 'r's or 'q's and 'g's should be easily confused, whereas 'n's and 'm's or even 'e's and 's's (which are syntactically similar but not similar at the role level) should not.
4. Errors should reflect the underlying role structure of letters. The more ambiguous a letterform's parts are, and the more tenuously they fill roles, the longer and less accurate the classification should be.

Method

Stimuli. A black grid was displayed against a white background on a Macintosh screen. The 21 vertices were drawn with a unit distance of 30 pixels (1.311 cm) making the entire grid 7.9×2.6 cm. A 3×3 pixel mark was made at each vertex. Every quantum in the grid was drawn in with a thickness of one pixel. Thus the grid provided noise for each gridfont letter (or, *token*) displayed. A token was displayed by drawing its active quanta with a thickness of three pixels. Subjects were seated, but able to maintain whatever distance from the monitor they found best.

Procedure. A set of gridfont characters was shown to each subject. The tokens were presented individually, with the screen blank for two seconds between each trial.

The full data set, **ALL**, may be meaningfully separated into two exclusive subsets: 1) **NORMALS** — 389 unique tokens which represent each of the 26 lowercase letter categories from 9 to 24 times and are intended to be very central examples of each letter; and, 2)**FONTS** — consisting of six entire gridfonts, each a coherent set of 26 tokens (a typeface) meant to share the same stylistic identity (or *spirit*), ranging from eccentric and style-saturated to reasonably normal.

Subjects were told that they would see a series of stimuli, each of which was intended to represent a roman lowercase letter. Responses were to be prompt keyboard keypresses which identified the stimulus as one of the 26 roman lowercase letters. The importance of correct identification was stressed

Letter Perception: Toward a conceptual approach

Gary McGraw & John Rehling
Center for Research on Concepts and Cognition
Indiana University, Bloomingon, Indiana 47405

Istituto per la Ricerca Scientifica e Tecnologica
Loc. Pantè di Povo, I38100 Trento, Italy
gem@irst.it rehling@irst.it

Robert Goldstone
Department of Psychology
Indiana University
Bloomington, Indiana 47405
rgoldsto@ucs.indiana.edu

Abstract

We present the results of a simple experiment in lowercase letter recognition. Unlike most psychology studies of letter recognition, we include in our data set letters at the extremes of their categories and investigate the recognition of letters of multiple typefaces. We are interested in the relationship between the recognition of normal letters and the recognition of non-standard letters. Results provide empirical evidence for top-down conceptual constraints on letter perception in the form of roles and relations between perceptually-based structural subcomponents. A process model based on the hypothesis developed below is currently being implemented.

Letters and Concepts

Literate humans are adept at recognizing letters in many diverse forms. Large variations in style — especially apparent in advertising faces — do not seem to hinder our ability to accurately identify characters. Character recognition by people is still vastly superior to that of machines. Why is it that people are able to quickly and easily recognize letterforms while machine recognition systems fail? We believe what is required is an approach using abstract conceptual representations of letters like those found in the human mind. We will describe such representations and empirically justify their existence.

Letterforms are subtler than people generally realize. Most people think the letter 'a' is just a *shape*. A closer look reveals an interconnected web of abstractions that make up the idea of 'a' itself. In one common form, the concept lowercase-'a' can be thought of as a marriage of two smaller ideas: (1) the idea of a curved umbrella-handle-like bar on the right, and (2) the idea of a small c-like curve nestled below the umbrella-handle. (See the first glyph of Figure 1.) These two conceptual components, which we call *roles*, are not explicit shapes *per se* but are ideas about what the acceptable bounds for letter-part shapes are, how far such shapes can be stretched before they lose their interpretation, and how they interact with other roles to form a complete object.[1]

Roles, like *wholes* (complete letters), are concepts in their own right, with somewhat nebulous boundaries. The difference is that membership in a role is easier to characterize than membership in a whole, so that reducing wholes to collections of interacting roles is a step forward in simplification.[2] (Palmer, 1978) points out the existence and utility of representations of structural units larger than lines and/or points in perceptual processing. Under our hypothesis, roles provide top-down influence on the parsing of figures into higher-order structures that have been shown to be important for making similarity judgements.

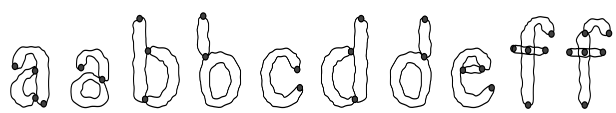

Figure 1: A picture of some of the most standard sets of roles for 6 lowercase letters. Roles are not themselves shapes, instead they are sets of norms at the conceptual level into which parts of letters can be fit. These pictures show how such roles work together. The black dots represent between-role relations.

Figure 1 is a graphical representation of some of the most standard sets of roles for the first 6 lowercase roman letters. Roles and relationships between roles make up the internal structure of a letter category. Category membership at the whole-letter level is determined by category membership at the lower level of roles. The style of a letterform is a function of how the various roles are filled.

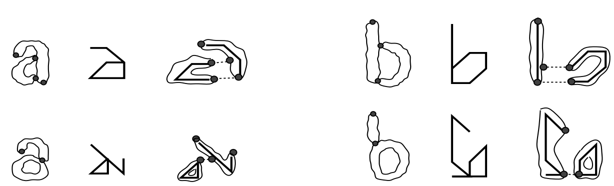

Figure 2: Parsing letterforms into high-level parts occurs under the top-down influence of roles. The top two examples show straightforward parsings (or, *role-fillings*) requiring little slippage. The bottom two are more complicated, with role-slippage performing a critical function in recognition.

[1](Blesser et al., 1973) discuss a related idea.

[2]The facilitating effect of similar perceptual hierarchies are discussed in (Palmer, 1977). Such hierarchies have rarely been applied to machine-based letter perception (Mantas, 1986).