Journal of
Experimental Psychology

VOL. 54, NO. 2 AUGUST, 1957

TRANSFER OF EXPERIENCE WITH A CLASS-SCHEMA TO IDENTIFICATION-LEARNING OF PATTERNS AND SHAPES [1]

FRED ATTNEAVE

Operator Laboratory, Air Force Personnel and Training Research Center

The idea that experience provides the organism with some "apperceptive mass" which facilitates later perception and learning is an old one, with many variants and elaborations. The concept of a "schema" as an entity mediating the effects of past experience has been prominent in recent British psychology, chiefly because of the thinking of Bartlett (**5**) and his associates, as reflected in the summary and review of Oldfield and Zangwill (**12**). Woodworth (**13**) and Hebb (**7**) have used the term "schema" in a sense which is somewhat more restricted and more definite than Bartlett's. After considering a number of experiments on memory for form, Woodworth concluded that a new configuration is usually remembered in terms of a "schema, with correction." For example, a figure which may be described as "a square with a nick on one side" is easier to learn than most other seven-sided polygons because the schema "square" is simple, familiar, and unambiguous, and the correction "with a nick in one side" is easily and clearly specifiable. Hebb emphasizes the importance of acquiring the schema of a class of objects to be differentiated from one another: thus Chinese "all look alike" to the occidental observer who has not seen many Chinese. So conceived, the schema consists, at least in part, of some representation of the central tendency or communality of the class of objects in question. If the observer has some subjective standard of the human face which he has obtained by "averaging" the faces of Americans, he may learn a new American face in terms of the manner and degree in which it deviates from this schema (cf. Woodworth's "correction"). If he is suddenly thrust into a Chinese population, however, his standard will no longer be central, and the new faces will all deviate from it in more or less the same direction.

Recently Oldfield (**11**), Hochberg (**8**), and the writer (**3**) have separately pointed out that the use of schemata makes for economical information storage. Oldfield's development of

---

[1] This report is based on work done under ARDC Project No. 7706, Task No. 27001, in support of the research and development program of the Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

this idea is the most detailed: Whatever suspicion one might have that the schema concept need be subjective and amorphous is removed by his demonstration that a computer might employ schemata (in essentially the Woodworth-Hebb sense) to good advantage.

In view of the theoretical attractiveness of this concept, it is surprising that little or no directly relevant experimentation has been undertaken. To the best of the writer's knowledge, the two experiments reported here constitute the first test under controlled conditions of the hypothesis, most directly attributable to Hebb, that the learning of a class schema makes easier the subsequent learning of identifying responses to members of the class.

EXPERIMENT I: LETTER–PATTERNS

Method

Materials.—The experimental stimuli consisted of three prototype patterns, and eight variations on each prototype. A *prototype* was constructed in the following manner. In a matrix six units wide and five high, each cell was either filled with a letter of the alphabet or left blank. For each cell, the probability of a blank was ½, and the probability of any particular letter was 1/52. The state of each cell was independently determined by the use of a table of random numbers, in accordance with these probabilities. A *variation* was a pattern in which one letter of the prototype was changed to some other letter (blanks were never varied). Both the position of the letter changed and the identity of the letter to which it was changed were randomly chosen. Figure 1 shows one of the three prototypes and four of the eight random variations on it. Note that the variations necessarily differ from one another more than they differ from the prototype.

```
  S   K Y G            S   K Y G
F   Y P Y K          F   Y P Z K
P     O F            P     O F
R   I   I            R   I   I
  D       W            D       P

  S   K Y G            A   K Y G
F   Y R Z K          F   Y P Z K
P     O F            P     O F
R   I   I            R   I   I
  D       W            D       W

            S   K Y G
          F   Y P Z K
          P     O F
          R   I   I
            D       W
```

FIG. 1. Letter patterns used in Exp. I. A prototype (bottom) and 4 of the 8 variations on it.

In the actual stimulus patterns, lines marking off rows and columns were omitted, as in Fig. 1. The patterns were typed with an electric typewriter (single spacing between rows; double spacing between columns) on a fine-grain, translucent tracing paper, behind which a reversed sheet of carbon paper was placed to increase the density of the typing. The patterns were then cut out and mounted in 2 × 2 in. slides, with due attention to uniformity of framing. Although the paper somewhat reduced the brightness of the ground, the projected images were quite satisfactory with respect to contrast and legibility.

Subjects.—A total of 60 airmen (basic trainees) at Lackland Air Force Base served as *S*s.

Procedure.—The 30 *S*s assigned to Group E were given pretraining which consisted of repeated viewings of one of the prototype patterns, interspersed with attempts to reproduce it, whereas the 30 *S*s of Group C were given an equivalent amount of practice in the reproduction of a completely irrelevant figure (a nonsense shape). Both groups were then tested on a paired-associates learning task in which the stimulus members were variations on the prototype used in the pretraining of Group E.

Pretraining consisted of eight reproduction trials. The pretraining stimulus was exposed (projected to the wall in front of *S*) for 15 sec. each time; after each exposure *S* attempted to put on paper as accurate a reproduction of the stimulus as he could. In the paired-associates learning task, the stimuli were eight variations on the prototype reproduced by Group E, and the responses were eight names of men (three letters each; no two with the same initial letter; e.g., *Sam*, *Joe*). Six learning trials were given, each followed by a test trial. Twelve different random permutations were used for these 12 presentations of the stimuli. All stimulus exposures were of 10 sec. duration, with 5 sec. between exposures on the same trial: the slide projector was automatically stepped at this rate by a set of Hunter timers. On learning trials, the "name" of each pattern was given aloud from a tape recording as the pattern was projected (the tape and the timing mechanism of the projector were synchronized by *E* at the beginning of the trial). On test trials, *S* attempted to write the correct name of each pattern, as it appeared, in a mimeographed booklet which provided an individual sheet for every trial. The interval between successive trials was 1 min., during which *E* rearranged the slides for the next presentation.

The *S*s were taken in subgroups of 10 at a time. Each of the three subgroups making up Group E was given a different prototype and associated variations. The same three sets of variations were likewise assigned to the three subgroups of Group C.

At no time was *S* told that the pretraining activity might improve his performance on a subsequent task; instead he was given the impression that the reproduction and the identification learning were two separate "memory tests." The only mention of a connection between the two was made to Group E, near the beginning of the instructions preceding paired-associates learning: "Now I'm going to give you a different kind of memory test. I'm going to show you eight patterns, made up of letters of the alphabet. They are all pretty much alike (in fact, they are all variations of the pattern you've been copying) but they are different enough so that you can learn to tell them apart . . . ." The parenthetical clause was of course omitted for Group C.

The objective of providing experimental *S*s with a subjective standard or schema corresponding to the prototype pattern might have been accomplished by various pretraining techniques other than reproduction. Arnoult (1) has found, however, that reproduction is the most effective of a wide variety of methods for learning a pattern, when the criterion of learning is the ability to recognize the pattern in a context of others very similar. Reproductions of the prototype by Group E were not systematically scored, but were inspected fairly thoroughly. Some *S*s achieved formally perfect reproductions; typically, however, a few errors were still present on the last trial.

Results

Errors made during paired-associates learning were summated over all six trials for each *S*; these total scores were then averaged within groups. The mean error score was 36.20 for Group C and 28.47—about 21% less—for Group E. It should be added, however, that the task was altogether too difficult for a good many of the *S*s: on the sixth and final trial, 18 *S*s (5 in Group E, 13 in Group C) were still making purely chance scores of 7 or 8 errors. On the same trial, 13 experimental *S*s and 6 control *S*s made no more than 2 errors; the modal numbers of errors was zero (with 9 *S*s) for Group E and 7 (with 8 *S*s) for Group C.

Since the results showed no evidence of any real differences between subgroups given different specific patterns, subgroups were simply pooled in a t test of the difference between total errors of Group E and Group C. A value of $t = 2.76$ was obtained; with $df = 58$, this is significant at the .01 level.

EXPERIMENT II: POLYGONS

In the case of the second experiment, which employed angular nonsense shapes instead of letter patterns, the advantages of pretraining with a prototype were somewhat less obvious, by intuitive standards, than formerly. Also the design was more complex, introducing an additional variable having to do with the way in which the variations differed from the prototype and from one another.

Method

Materials.—Ten prototype shapes were constructed by plotting points with random coordinates in a 16 × 16 matrix, and then connecting the points with the shortest possible closed contour (4). Five 6-sided and five 12-sided polygons were thus prepared: all 10 are reproduced in the two upper rows of Fig. 2. In each of the variations on such a prototype, one-third of its points (2 out of 6, or 4 out of 12) were moved. The movement of a point was from its original matrix cell to a randomly chosen one of the eight immediately surrounding cells: i.e., the direction of movement varied randomly, but the extent of movement was approximately constant (either 1 or $\sqrt{2}$ matrix units). From each prototype, two sets of eight variations were made. In the first set, the *same* two or four points were varied throughout the whole set (though the selection of these points was initially random). In the

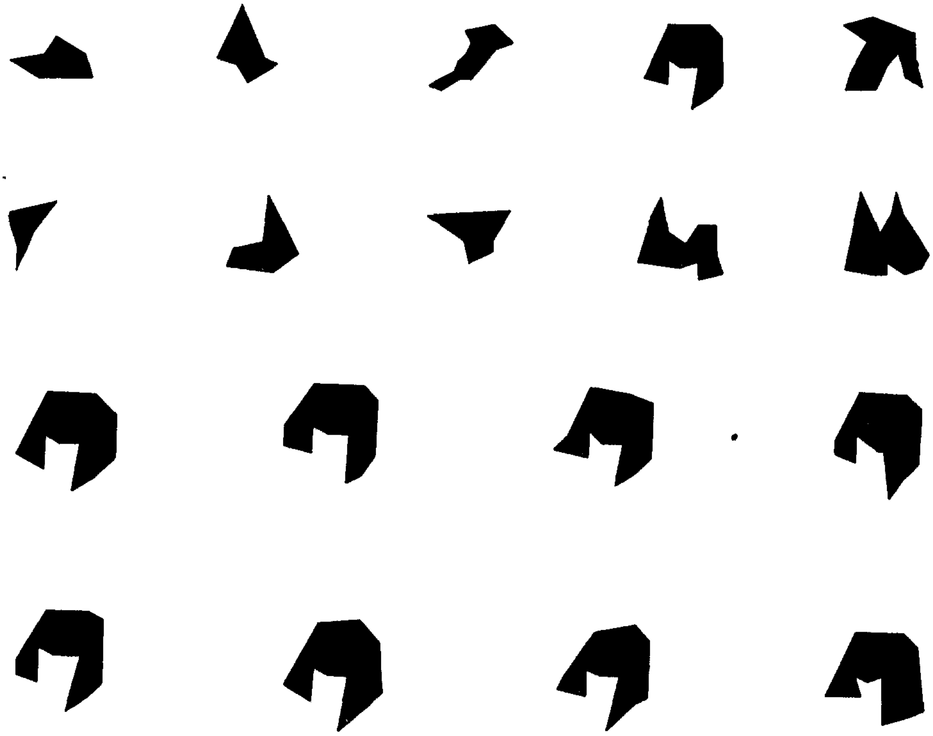

FIG. 2. The 10 shapes used as prototypes in Exp. II are shown in the two upper rows; a sample set of variations is shown in the two lower rows.

second, an independent random selection of the two or four points to be moved was made for each individual variation. In other words, the *loci* of variation were constant for the first set and randomly variable for the second.

The two lower rows of Fig. 2 show a set of variations on one of the prototypes. It follows from the method of construction that the variations "vary about" the prototype: i.e., that the prototype represents the central tendency of the class.

Each of the 170 shapes (10 prototypes and 20 sets of 8 variations) was mounted in a 2 × 2 in. slide. The figure-ground relationship was the reverse of that shown in Fig. 2, i.e., the shape was transparent and the remainder of the slide opaque. No matrix lines appeared on the slide.

Subjects.—The Ss were 320 airmen from the same population as in Exp. I.

Procedure.—In all essential respects, the procedure was the same as in Exp. I. Subgroups were made up of 8 instead of 10 Ss, an experimental and a control subgroup being employed with each of the 20 sets of variations. Experimental Ss were again given pretraining which consisted of eight reproduction trials with the relevant prototype. Control Ss were given pretraining with an unrelated shape: if their paired-associates stimuli were to be 6-sided polygons, they were pretrained with one of the 12-sided polygons, and vice versa. The first eight letters of the alphabet, instead of men's names, were used as response members of the paired associates. Number and duration of trials were as before.

Results

Differences in errors on the paired-associates task appeared on first analysis to be small and variable: 7 of the 20 matched pairs of subgroups showed differences in the unexpected direction (see Table 1). With the view of compensating for the high degree of inter-*S* variability, a predictor variable which might be used for covariance control was sought. The

TABLE 1

SCORES ON EXPERIMENTAL TASK AND ON TEST OF INITIAL ABILITY

Pairs of Subgroups	Mean Error		Test Score		r (Error-ML)	Unadjusted Error Diff.	Adjusted Error Diff.
	Exp.	Control	Exp.	Control			
6Sa	18.13	24.25	4.38	4.63	−.61	6.12	6.56
6Ra	28.25	29.75	3.13	4.38	−.72	1.50	6.42
6Sb	33.63	31.75	3.25	5.38	−.60	−1.87	3.44
6Rb	26.25	26.88	4.75	4.88	−.53	.62	.86
6Sc	35.88	34.00	3.75	4.25	−.42	−1.87	−1.16
6Rc	29.63	27.00	3.88	4.63	−.33	−2.62	−1.42
6Sd	26.38	29.88	3.88	4.38	−.73	3.50	5.09
6Rd	31.38	32.75	4.88	4.38	−.41	1.37	.58
6Se	24.38	27.75	4.88	4.38	−.58	3.37	2.31
6Re	27.63	26.25	4.63	3.88	+.01	−1.37	−1.34
12Sa	25.00	22.50	3.50	5.50	−.45	−2.50	3.56
12Ra	27.13	28.62	3.13	3.25	−.56	1.50	1.94
12Sb	28.00	30.38	4.75	5.00	−.56	2.37	3.48
12Rb	29.25	23.13	4.50	6.75	−.39	−6.12	−2.62
12Sc	27.50	30.50	3.88	4.75	−.83	3.00	6.53
12Rc	20.50	30.63	5.63	4.63	−.41	10.12	7.41
12Sd	28.13	29.25	2.75	4.13	−.47	1.12	4.54
12Rd	23.00	27.00	5.75	5.13	−.17	4.00	3.49
12Se	27.13	28.13	3.88	5.00	−.32	1.00	2.84
12Re	26.13	20.88	3.25	5.00	−.54	−5.25	.97
Mean	27.17	28.06	4.12	4.72	−0.52*	.90	2.67

* Root-mean square.

classification battery given to all Air Force basic trainees contains a test known as "Memory for Landmarks" which is not unlike the identification-learning task of the present experiment, and which showed a correlation with the present task of $r = -.52$ (RMS average of within-subgroups correlations). Moreover, control *S*s were found to have been disconcertingly higher in initial ability, by the "Memory for Landmarks" criterion, than experimental *S*s (see Table 1); thus, some control for initial ability was even more strongly indicated.

Now, the experiment was originally so designed that an error term might be based upon variability among subgroups differing in both *S*s and stimuli, in order that "significant" results might be considered to hold for a parent population of stimuli, as well as for a parent population of *S*s (4). By adhering to this strict criterion of significance, and excluding from consideration any weaker types of generalization, it was possible to handle the covariance adjustment and the subsequent tests in an unusually simple and economical manner.

The data were first simplified by obtaining difference scores for matched *E* and *C* subgroups (see Table 1). An adjustment for initial ability was then made independently on each of these 20 difference scores, by the formula

$$\text{Adjusted } D_x = D_x - b_{xy} D_y,$$

in which D_x is the difference between the subgroup means on the experimental task, D_y is the difference on the "Memory for Landmarks" test, and b_{xy} is the within-subgroups regression coefficient. It should be emphasized that b_{xy} was obtained separately for

TABLE 2

Error–Differences Between Covariance-Adjusted Means of Matched Experimental and Control Subgroups

Sides	Prototype	Points Varied		S+R	S−R
		S(ame)	R(andom)		
6	6a	6.56	6.42	12.98	.14
	6b	3.44	.86	4.30	2.58
	6c	−1.16	−1.42	−2.58	.26
	6d	5.09	.58	5.67	4.51
	6e	2.31	−1.34	.97	3.65
12	12a	3.56	1.94	5.50	1.62
	12b	3.48	−2.62	.86	6.10
	12c	6.53	7.41	13.94	−.88
	12d	4.54	3.49	8.03	1.05
	12e	2.84	.97	3.81	1.87
Mean		3.72	1.63	5.35	2.09
t ($df = 9$)		5.25	1.56	3.25	3.06
P (2-tail test)		<.001	<.20	=.01	<.02

Note:—In Columns S, R, and S+R, a negative sign indicates that control *S*s made fewer errors than experimental *S*s. Each *t* is a test of the null hypothesis that the column mean is a chance deviation from zero.

each pair of subgroups to allow for the possibility of true differences in correlation dependent on stimulus materials; also because it was considered desirable that the adjusted difference scores be completely independent of one another with respect to statistical restraints. Table 1 shows correlation coefficients rather than regression coefficients; the relation between the two is: $b_{xy} = r \frac{\sigma_x}{\sigma_y}$ (10, Ch. 15).

The adjusted difference scores are classified by systematic stimulus characteristics in Table 2. These 20 scores are still not entirely independent with respect to stimuli, because two sets of variations (one with the same points moved, the other with random points moved) were derived from each of the 10 prototypes. A set of 10 completely independent scores (the "S + R" column of Table 2) was obtained by combining difference scores associated with the same prototype. A *t* test of the departure of this distribution from zero yields a value of 3.25, with $df = 9$, for which P is just equal to .01. This measures our confidence that the over-all difference between Group E and Group C is generalizable both to other *S*s and to other stimuli constructed by the same rules.

Considering results on "Same" and "Random" variations separately, in the former case a *t* of 5.25 is found with $df = 9$, significant at the .001 level, and in the latter case there is a nonsignificant *t* of 1.56 in the expected direction. The "S − R" column of Table 1 contains the distribution of differences between difference scores obtained with "Same" and "Random" variations, matched by prototype. The *t* for the departure of this distribution from zero is 3.06 with $df = 9$, significant at the .02 level. One may be reasonably sure, therefore, that pretraining with the prototype did more good when the same parts of the figure were subsequently varied than when different parts were varied. A plausible interpretation of this result, in crude terms, is as follows: When an *S* who has become familiar with the prototype sees the first of the variations, his attention is drawn to those parts of the figure which differ from the prototype, and a set with respect to *where to look* for distinguishing characteristics is thereby established. This set is appropriate and useful if the loci of variation are constant throughout the class, but not otherwise. One would be ill-advised, however, actually to accept the null hypothesis associated with the nonsignificant *t* for the "Random" condition, and to conclude that familiarity with a prototype is effective *only* when the loci of variation are constant, for at least two reasons: First, and most important, positive results were obtained in Exp. I, in which locus of variation was not constant. Second, the actual odds against obtaining by chance a t as great as the one in question ($t = 1.56$, $df = 9$) are about 6 to 1, and conventions having to do with "significance" need not prevent us from giving these odds their due weight.

The experimental effect did not appear to vary with stimulus complexity: a comparison of "S + R" values for 6- vs. 12-sided polygons yielded a t of only .37 with $df = 8$. A similar test on the "S − R" column (essentially a test for interaction) likewise yielded a nonsignificant $t = .06$.

Discussion

Both of the experiments reported involve a possible attenuation of effects which should be considered carefully in the evaluation of the results. It may not be assumed that control *S*s learned to identify the variations without the benefit of a class schema, since there was nothing to prevent them from abstracting the schema once they were exposed to the class. Indeed, the very principle on which one would predict superiority of Group E over Group C would normally apply to situations, not unlike that confronted by the control *S*s, in which schemata are available only by abstraction from natural objects. The most one can assume is that the pretraining of experimental *S*s gave them a head start over the control *S*s in the acquisition of an appropriate schema. Accordingly, the obtained differences in favor of Group E can be considered no more than dilute manifestations of the effect of schemata on identification learning.

It is evident that a close relationship must exist between the mechanisms associated with schemata and with "stimulus-predifferentiation," as studied by the Gibsons and others (2). The present experiments do not involve predifferentiation in any literal sense, for the pretraining was on a single stimulus; on the other hand one may justifiably assume that predifferentiation training does involve schema learning. This is not to say that the predifferentiation subject acquires nothing more than a knowledge of the "average" stimulus. More likely, he learns something about at least three characteristics of the class: (*a*) its central tendency; (*b*) *how* its members may differ from one another, i.e., in what properties, or on what dimensions; and (*c*) its dispersion, i.e., *how much* its members may differ from one another on the several dimensions of variability. The importance of (*a*) is of course indicated in the present studies. Sensitivity to both (*a*) and (*c*), in unidimensional situations, is implied by the relativity of judgment which individuals commonly display in assigning stimuli to rating-scale categories. The importance which (*b*) assumes in multivariate situations has recently been shown by Kurtz (9) (cf. also the difference between the "Same" and "Random" conditions in Exp. II). Although in the present paper *schema* has been used in the Woodworth-Hebb sense, to refer to some representation of the central tendency of a multivariate class, there is perhaps no reason why its reference should not be extended to include the whole system of class parameters suggested above.

The importance of schemata in learning is supported by so much commonsense evidence that the present studies may well be considered demonstrations rather than experiments. It is hoped, however, that these studies may open the way for others investigating more uncertain matters. A particularly interesting question concerns the effect of pretraining with a noncentral standard, i.e., one which is similar but peripheral to the stimuli of the paired-associates class. Positive transfer might be predicted on the ground that the standard shares information with the paired-associates stimuli; alternatively, negative transfer might be predicted on the ground that the stimuli to be identified all deviate in the same direction from the standard, and therefore should appear more alike.

The answer to this question would throw some light on the mechanisms underlying the present findings.

SUMMARY

Two experiments were conducted to determine the effect on a paired-associates learning task of prior familiarization with a single "prototype" stimulus representing the central tendency of the stimuli to be identified. The stimuli were letter patterns in Exp. I, and polygons in Exp. II. In both cases they were constructed by systematically random methods.

In Exp. II a sufficient sample of shapes was used to permit statistical generalization of the results to other stimuli as well as to other *S*s. Two different methods were employed in this experiment to generate the paired-associates stimuli: in one, the same parts of the shapes were varied throughout the set; in the other, different parts were varied from shape to shape.

Positive results were obtained in both experiments: i.e., familiarization with the central "prototype" stimulus decreased errors on the paired-associates task. In Exp. II this effect was greater when the loci of variation among the shapes were constant.

REFERENCES

1. ARNOULT, M. D. A comparison of training methods in the recognition of spatial patterns. Lackland AFB, Tex.: Air Force Personnel and Training Research Center, February, 1956. (*Res. Rep.* AFPTRC-TN-56-27.)
2. ARNOULT, M. D. Stimulus predifferentiation: some generalizations and hypotheses. *Psychol. Bull.*, 1957, **54**, 340–351.
3. ATTNEAVE, F. Some informational aspects of visual perception. *Psychol. Rev.*, 1954, **61**, 183–193.
4. ATTNEAVE, F., & ARNOULT, M. D. The quantitative study of shape and pattern perception. *Psychol. Bull.*, 1956, **53**, 452–471.
5. BARTLETT, F. C. *Remembering, a study in experimental and social psychology.* Cambridge: Cambridge Univer. Press, 1932.
6. GUILFORD, J. P. (Ed.) *Printed classification tests.* Washington: U. S. Govt. Print. Off., 1947. Army Air Forces Aviation Psychology Program Res. Rep. No. 5.
7. HEBB, D. O. *Organization of behavior.* New York: Wiley, 1949.
8. HOCHBERG, J. E. Psychophysics and sterotypy in social perception. In *Emergent problems in social psychology* (in press).
9. KURTZ, K. H. Discrimination of complex stimuli: the relationship of training and test stimuli in transfer of discrimination. *J. exp. Psychol.*, 1955, **50**, 283–292.
10. MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1949.
11. OLDFIELD, R. C. Memory mechanisms and the theory of schemata. *Brit. J. Psychol.*, 1954, **45**, 14–23.
12. OLDFIELD, R. C., & ZANGWILL, O. L. Head's concept of the schema and its application in contemporary British psychology. *Brit. J. Psychol.*, 1942, **32**, 267–286; 1943, **33**, 58–64, 113–129, and 143–149.
13. WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938.

(Received July 3, 1956)