

Further studies on the spatial compatibility of four control–display linkages

Swei-Pi Wu *

Department of Industrial Management, Hua Fan College of Humanities and Technology, 1 Hua Fan Road, Shihtin Hsiang, Taipei Hsien 223, Taiwan, ROC

Received 21 February 1995; accepted 12 February 1996

Abstract

This study used a paper–pencil test and a physical model experiment to evaluate the suggestive effect, population stereotypes, and the preferred arrangement concerning the control–display relationship of a four-burner stove. A total of 1163 subjects (680 males and 483 females) was surveyed using four different questionnaire forms including alphabetical and sign code systems. The results show that the suggestive effect of the sequential codes indeed exists, and that the most frequently chosen arrangement by the Chinese subjects was Type III which is different from Type IV chosen by the American subjects. In addition, thirty male subjects carried out trials on a wooden physical model of the four-burner stove and their reaction times and error rates were measured. The results reveal that Type V was the preferred arrangement for the Chinese subjects and Type II for the American subjects. In general, this study verified that neither the paper–pencil test nor the computer simulation research methods are equivalent to the physical model experiment. It is probably best to use actual performance on a physical model to investigate the compatibility or population stereotypes between controls and displays.

Relevance to industry

Compatibility is a central concept in human factors. A major goal in any design is to make the system compatible with human expectations. From such a viewpoint, how to investigate the compatibility correctly is very important. Additionally, the compatibility difference between Oriental and Occidental peoples should be investigated.

Keywords: Compatibility; Control–display relationship; Population stereotype; Methodology

1. Introduction

Besides enhancing performance, reducing human error and accidents and improving safety remain priorities in any human factors effort. One of the

most effective human factors principles to obtain these objectives is to design a system compatible with human expectations or population stereotypes. Compatibility deals with the relationship of stimuli and responses to human expectations. In general, a

* Corresponding author. Tel.: +886-2-6632102; fax: +886-2-6631119.

system with a greater degree of compatibility will result in faster learning, faster response times, fewer errors, and a reduced mental workload. Although people can do learn to use systems that are out of sync with their expectations, they do so at a price (Sanders and McCormick, 1992).

In a classic demonstration of this principle, several studies have investigated the physical arrangement of the burner controls on a four-burner stove (Chapanis and Lindenbaum, 1959; Ray and Ray, 1979; Shinar and Acton, 1978; Osborne and Ellingstad, 1987; Hsu and Peng, 1993). Chapanis and Lindenbaum (1959) and Ray and Ray (1979) presented various arrangements of controls and burners (as shown in Fig. 1) to subjects and asked them to turn on specific burners. The number of errors was recorded. The results from both studies showed that, when burners are aligned (designed in quadrature), the arrangement of Type II appears to be superior.

Shinar and Acton (1978) conducted a survey to test the existence of population stereotypes concerning burner controls. They presented to subjects a questionnaire containing a drawing and asked them to indicate which of the unmarked controls they thought controlled each of the burners. The burners were labeled 'A', 'B', 'C', and 'D', and the controls were unlabeled. The results revealed at least four commonly expected different linkage relationships (including Type II, III, IV, and V), and with no one particular stereotype emerging as predominant among

them. Noteworthy, the most frequently chosen arrangement was Type III, being chosen by 31% of the subjects. Type II, which resulted in fewer errors than Type III, was chosen by only 25% of the subjects. It was inconsistent with the previous studies.

Hsu and Peng (1993) supposed the aforementioned discrepancy might have resulted from the 'suggestive effect' of the code system used in Shinar and Acton's study. Because the alphabetic codes A, B, C, and D are essentially sequential, some subjects might have been influenced unknowingly by this and filled in the four controls with those letters consecutively from left to right, thus facilitating a quasi-prevalence of the Type III choice. For this reason, Hsu and Peng developed three different questionnaire forms, including alphabetical, sign, and numerical code systems to evaluate the possible suggestive effect of sequential codes. The results indicated the existence of a suggestive effect, which is a tendency by some stimuli to induce a specific response unawares in subjects. In addition, Hsu and Peng (1993) also developed a computer simulation to model the four types (II, III, IV, and V) of burner controls linkage. The results revealed the population stereotype was different between Chinese subjects and American subjects. Type III represented the Chinese stereotype, and Type II the American stereotype.

After reviewing the literature of burner controls studies, we proposed the following issues: (1) Do the arrangements people choose always result in the optimum levels of performance? (2) Does the suggestive effect indeed exist, and does it significantly influence the arrangements people choose? (3) Are both the paper-pencil test and computer simulation research methods equivalent to the physical model?

Clearly, the above controversy must be resolved. The primary objective of the present investigation was to verify the existence of a suggestive effect when sequential codes are used in the questionnaire. A secondary objective of this study was to evaluate the equivalence of research methods among the paper-pencil test, computer simulation and physical model. The last objective of this research was to know whether or not a difference in the population stereotype existed between Chinese and American subjects regarding the control-burner relationship. Based upon the point of view of ergonomics, it is believed that these data would be useful in providing

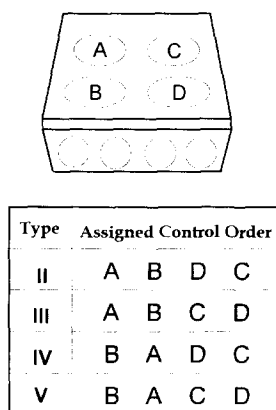


Fig. 1. Control-burner arrangement relationship used in previous studies.

a guide for designing a system regarding to control–display relationship.

2. Methods

2.1. Paper–pencil test

2.1.1. Subjects

A total of 1163 subjects (680 males and 483 females) participated in this study. They were paid volunteer college students. Subjects were recruited from six colleges, their ages ranged from 15 to 22 years.

2.1.2. Survey questionnaires

The questionnaires used in the study were revised from Shinar and Acton (1978) and Hsu and Peng (1993). In order to evaluate the suggestive effect of sequential codes effectively, four different forms of questionnaires were developed. Noteworthily, only two code systems – alphabetical and sign – were employed in these four questionnaires (named Qa, Qa', Qs, Qs') and used in models similar to Fig. 2.

In the Qa and Qs questionnaires, the alphabetic codes (A, B, C, and D) and sign codes (☆, #, △, and □) were labeled over the four burners, respectively. For the purpose of testing the suggestive effect of these two code systems, the alphabetic codes (A, B, C, and D) and sign codes (☆, #, △, and □) were

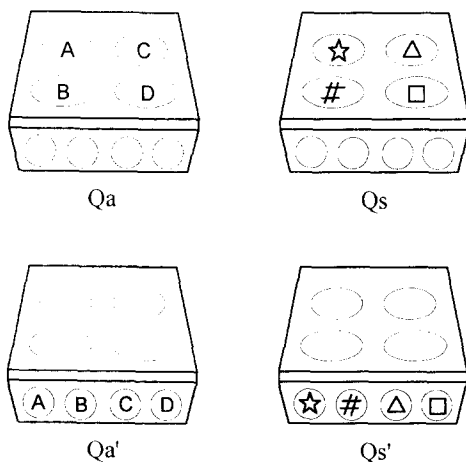


Fig. 2. The four forms of questionnaires used in this study.



Fig. 3. A wooden model of the four-burner stove used in this study.

allocated to each control from left to right in the other questionnaires Qa' and Qs', respectively.

2.1.3. Data collection

The four types of questionnaires were randomly distributed to the 1163 subjects: Qa went to 302 subjects (206 males and 96 females), Qa' to 268 subjects (164 males and 104 females), Qs to 347 subjects (170 males and 177 females), and Qs' to 246 subjects (140 males and 106 females).

At the beginning of the data collection, subjects were briefed about the purpose of the survey. In addition, each subject was instructed to fill the questionnaires intuitively according to his or her stereotyped judgement. For Qa and Qs questionnaires, the subject's task was to fill in each control blank with its corresponding burner's code, and conversely to fill in each burner blank with its corresponding control's code for Qa' and Qs' questionnaires.

2.2. Physical model experiment

2.2.1. Subjects

Thirty paid volunteer college male students were recruited for this study. The subjects' ages ranged from 18 to 22 years, with the average age being 19.8 years (SD = 0.805 years). The subjects were familiarized with the experimental purpose and procedure before the experimental data were collected.

2.2.2. Apparatus

The apparatus used in this experiment consists of a wooden model of a four-burner stove and a com-

puter as shown in Fig. 3. Based upon previous studies (Chapanis and Lindenbaum, 1959; Ray and Ray, 1979), a four-burner stove made of wood was modified. The dimensions of this wooden model were 46 cm wide, 61 cm long, and 15 cm high. The burners, represented by colored perspex discs in the center of each of which was a small light (made of a LED), were aligned on the horizontal surface of the wooden model. The controls, represented by push button switches, were fitted on the front vertical panel of the burners. The circuits were arranged so that any one of the burners could be lit automatically by the computer and turned off by the subject pushing the correct control knob. In addition, a timer in the computer connected with the circuits served as the measuring instrument to record the reaction times and the number of errors.

2.2.3. Experimental design

A randomized complete block design with subjects as blocks was used to study the effect of control–burner linkage on the reaction times and number of errors. Four different types of control–burner linkage were investigated; they were Type II, III, IV, and V. Each subject carried out trials for all four different types in a random order. Because linkage type was varied within subjects, a two-day break was instituted between testing sessions on the different types to minimize any effect of learning.

2.2.4. Experimental procedure

The experimental procedure was similar to that of Chapanis and Lindenbaum (1959) and Hsu and Peng (1993). At the beginning of a test session the experimenter explained to each subject the type of control–burner linkage that would be used. The subject was allowed to adjust his seat to a comfortable height, and was asked to start with the index fingers of his preferred hand on a starting point 15 cm away from the center of the four controls. When the subject was ready, the experimenter would give a ready signal and, shortly thereafter, turn on a stimulus light in one of the four burners. The subjects' task was to turn off the light by pushing the correct control button as quickly as possible. If the subject made an error, he had to immediately try another control button until the right one was found. Before the test began, the subjects was allowed practice ten

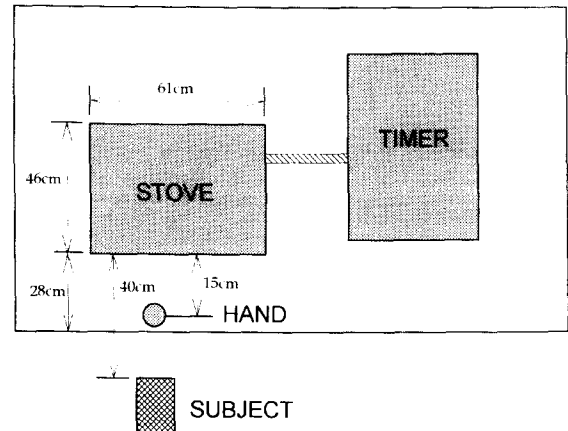


Fig. 4. Demonstration of the workplace layout for the experiment.

trials to make sure that they were familiar with the experimental task and control–burner relationship. The subject was also allowed to ask any question he might have about his task. After these were resolved the subject was tested for 80 consecutive trials, while each light burner was turned on and off 20 times randomly. The intertrial interval between light off and on was fixed at about 1 s. The criterion measures in this experiment were reaction times and number of errors, which were automatically recorded by the timer in the computer. Fig. 4 demonstrates the workplace layout of the experiment.

3. Results and discussion

3.1. Paper–pencil test

Table 1 contains the numbers of intuitive choices and the percentage of each possible control–burner linkage arrangement made by subjects in the four forms of questionnaires. The different control–burner linkages are coded according to the sequence of the control knobs activating burners A, B, C, and D in the Qa questionnaire.

3.1.1. Suggestive effect

As expected, Table 1 shows that the arrangement of Type III was chosen most frequently in all four questionnaires, and that Type III was chosen by a markedly high percentage of subjects in the case of

Table 1
Results of the pencil-paper test

Linkage type	relationship	Qa (N = 302)			Qa' (N = 208)			Qs (N = 347)			Qs' (N = 246)		
		male	female	total	male	female	total	male	female	total	male	female	total
II	ABDC ^a	57 (27.7) ^b	5 (5.2)	62 (20.5)	44 (26.8)	13 (12.5)	57 (21.3)	39 (22.9)	23 (13.0)	62 (17.9)	39 (27.9)	17 (16.0)	56 (22.8)
III	ABCD	87 (42.2)	68 (70.8)	155 (51.3)	56 (34.1)	28 (26.9)	84 (31.3)	64 (37.6)	59 (33.3)	123 (35.4)	46 (32.9)	43 (40.6)	89 (36.2)
IV	BADC	23 (11.2)	3 (3.1)	26 (8.6)	22 (13.4)	15 (14.4)	37 (13.8)	30 (17.6)	17 (9.6)	47 (13.5)	17 (12.1)	16 (15.1)	33 (13.4)
V	BACD	16 (7.8)	2 (2.1)	18 (6.0)	15 (9.1)	16 (15.4)	31 (11.6)	20 (11.8)	16 (9.0)	36 (10.4)	16 (11.4)	6 (5.7)	22 (8.9)
VI	BDAC	—	—	—	—	—	—	8 (4.7)	13 (7.3)	21 (6.1)	—	—	—
VII	ACBD	12 (5.8)	9 (9.4)	21 (7.0)	12 (7.3)	12 (11.5)	24 (9.0)	5 (2.9)	27 (15.3)	32 (9.2)	6 (4.3)	9 (8.5)	15 (6.1)
VIII	ADBC	1 (0.5)	—	1 (0.3)	—	2 (1.9)	2 (0.7)	1 (0.6)	—	1 (0.3)	—	—	—
IX	BDCA	6 (2.9)	5 (5.2)	11 (3.6)	—	—	—	1 (0.6)	2 (1.1)	3 (0.9)	—	1 (0.9)	1 (0.4)
Others		4 (2.0)	4 (4.2)	8 (2.6)	15 (9.1)	18 (17.3)	33 (12.4)	2 (1.2)	20 (11.5)	22 (6.5)	16 (11.4)	14 (13.3)	30 (12.3)
Total		206	96	302	164	104	268	170	177	347	140	106	246

^a 'ABCD' represents '☆# △ □' in Qs and Qs'.

^b Numbers in parentheses are percentages.

Qa (51.3%) and a somewhat smaller percentage in the case of Qa' (31.3%), Qs (35.4%), and Qs' (36.2%). In order to test the suggestive effect of the alphabetic codes and sign codes used in the present study, the data of Type II, III, IV, and V were statistically analyzed using the chi-square test method. For the alphabetical questionnaires, the result showed that there was a significant difference between Qa and Qa' ($\chi^2(3) = 20.638$, $p < 0.01$). However, there was no significant difference between the sign questionnaires Qs and Qs' ($\chi^2(3) = 1.624$, $p > 0.05$). A further test also showed that there was no significant difference among Qa', Qs, and Qs' questionnaires ($\chi^2(6) = 3.17$, $p > 0.05$). In addition, Type III was chosen by subjects significantly more often in Qa than in Qa' ($\chi^2(1) = 6.30$, $p < 0.025$), Qs ($\chi^2(1) = 16.37$, $p < 0.001$), and Qs' ($\chi^2(1) = 12.61$, $p < 0.001$). Aforementioned results reveal that the sequential nature of the alphabetical codes has a suggestive effect on the choice of Type III in the Qa questionnaire, and the sign codes, devoid of sequentiality, were immune to this suggestive effect. These results were the same as those in the study of Hsu and Peng (1993).

3.1.2. Population stereotypes

Because the alphabetical codes A, B, C, and D significantly influenced the choice of Type III in the Qa questionnaire, the data of Qs were used to analyze the population stereotype of control–burner arrangement for Chinese subjects. As reported by Shinar and Acton (1978), four common expected differ-

ent types (II, III, IV, and V) were indicated by the subjects (see Table 1, Qs). A chi-square analysis showed that there were significant differences among Type III, Type II, Type IV, and Type V for male, female, and grand total, respectively ($\chi_m^2(3) = 27.8$, $\chi_f^2(3) = 43.4$, $\chi_t^2(3) = 67.5$; all values of $p < 0.001$). Furthermore, Type III was chosen significantly more often than Type II, Type IV, and Type V ($p < 0.001$), which did not differ significantly among themselves. Noteworthy, no significant difference was found between male and female subjects ($\chi^2(3) = 3.03$, $p > 0.05$). Consequently, Type III represents the population stereotype of control–burner arrangements for Chinese subjects. This is in agreement with the finding in the study of Hsu and Peng (1993) (see Table 2).

3.1.3. Comparison with past studies

Table 2 shows a comparison of population stereotypes from the present and the past studies (Hsu and Peng, 1993; Shinar and Acton, 1978). As can be seen the population stereotypes from the present study are consistent with the results of Hsu and Peng (1993) for the sign questionnaire (Qs). Table 2 also shows the change percentage, i.e., the amount of the suggestive effect for the Qa questionnaire. For example, the suggestive effect of Qa on Type III was 16% and 11% for the present study and that of Hsu and Peng (1993), respectively. If the suggestive effect was subtracted from the Shinar and Acton data, then Type IV would surpass Type II, Type III, and Type V (see Table 2), and represent the population stereo-

Table 2
Comparison of population stereotypes with previous studies

Type	Linkage relationship	This study			Hsu and Peng (1993)			Shinar and Acton (1978)	
		Qa	Qs	change	Qa	Qs	change	Qa	Qs
II	ABDC	20%	18%	+2%	16%	20%	-4%	25%	(23–29%)
III	ABCD	51%	35%	+16%	49%	38%	+11%	31%	(15–20%)
IV	BADC	9%	14%	-5%	14%	19%	-5%	28%	(33%)
V	BACD	6%	10%	-4%	9%	9%	0%	15%	(15–19%)
VI	BDAC	0%	6%	-6%	5%	4%	+1%	1%	(0–7%)
VII	ACBD	7%	9%	-2%	0%	7%	-7%	<1%	(2–7%)
VIII	ADBC	<1%	<1%	0%	0%	<1%	<-1%	<1%	(<1%)
IX	BDCA	4%	1%	+3%	0%	0%	0%	0%	(0%)
Others		3%	6%	-3%	7%	2%	+5%	0%	(0–3%)

Qa: Alphabetical questionnaire, Qs: Sign questionnaire.
Numbers in parentheses are predicted values of Qs.

Table 3
Results of the physical model experiment

Type	Linkage relationship	Average reaction time (s)	Number of errors (%)
II	ABDC	0.699	61 (2.5%)
III	ABCD	0.693	80 (3.3%)
IV	BADC	0.704	89 (3.7%)
V	BACD	0.637	49 (2.0%)

type for American subjects. This finding was not the same as that of Hsu and Peng (Hsu and Peng, 1993).

3.2. Physical model experiment

3.2.1. Preferred Arrangement

The mean values of the performance measures in reaction times and error rates for all subjects are presented in Table 3. With regard to reaction times, Fig. 5 shows that Type V was the best (0.637 s), followed by Type III (0.693 s) and Type II (0.699 s), while Type IV (0.704 s) was the worst. As for error rates, Type V was also the best (2.0%), Type II (2.5%) and Type III (3.3%) were next, and Type IV (3.7%) was the worst.

An analysis of variance showed that there were significant differences in the mean reaction times among the different types ($F(3, 87) = 2.05$, $p < 0.01$) and among the subjects ($F(29, 87) = 3.77$, $p < 0.05$). The Duncan multiple range test showed that the mean for Type V was significantly lower than those for Type III, II, and IV ($p < 0.05$), while no significant difference was found between Type III, II, and IV. In addition, a chi-square analysis shows that there was a significant difference for the number of errors among the types ($\chi^2(3) = 14.09$,

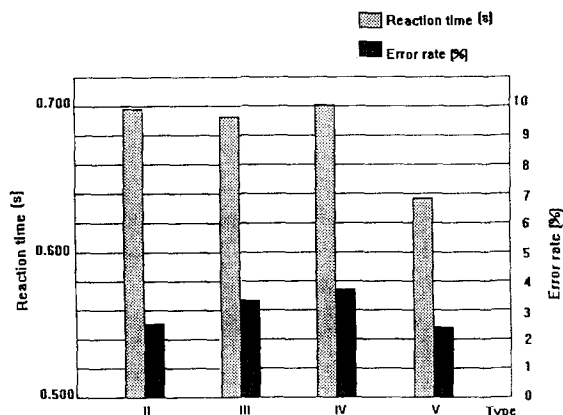


Fig. 5. The effect of different types of control–burner linkages on the reaction times and error rate.

$p < 0.005$); that of Type V was significantly lower than Type III and IV, but it did not differ significantly from that of Type II. Overall, Type V was confirmed as the preferred arrangement for the Chinese subjects. These results were not consistent with those of the paper–pencil test.

3.2.2. Comparison with past studies

Table 4 shows a comparison of the performance results from the present study and the past studies (Hsu and Peng, 1993; Chapanis and Lindenbaum, 1959; Ray and Ray, 1979). It is of importance to note that the preferred arrangement for this study was different from Hsu and Peng's computer simulation for Chinese subjects. However, the result of Ray and Ray (1979) was the same as that of Chapanis and Lindenbaum (1959) for American subjects. Obviously, the difference between the present and the Hsu and Peng (1993) study may be attributed to the fact that both the controls and burners of the com-

Table 4
Comparison of reaction time and error rate (%) with previous studies

Type	Linkage relationship	This study		Hsu and Peng (1993)		Chapanis and Lindenbaum (1959)	Ray and Ray (1979)
		reaction time	error rate (%)	reaction time	error rate (%)	error rate (%)	error rate (%)
II	ABDC	0.699	2.5%	0.665	11%	6% ^a	9% ^a
III	ABCD	0.693	3.3%	0.631 ^a	4% ^a	10%	16%
IV	BADC	0.704	3.7%	0.715	10%	11%	19%
V	BACD	0.637 ^a	2.0% ^a	0.720	11%		12%

^a The preferred arrangement for the study.

puter simulation and the paper–pencil test were in the same plane. Therefore, the physical model experiment should not be replaced, methodologically, by the computer simulation or paper–pencil test. According to this finding, the results of Shinar and Acton (1978) should not be consistent with those of the previous studies of Chapanis and Lindenbaum (1959) and Ray and Ray (1979). In general, it is probably best to use actual performance of the physical model to decide on arrangements between controls and displays.

4. Conclusions and recommendations

This study used two research methods, the paper–pencil test and the physical model experiment, to study the control–burner linkage relationship of the four-burner stove. The following conclusions are based on the results of these two methods, respectively.

4.1. Paper–pencil test

1. The sequential nature of the alphabetical codes, A, B, C, and D has a suggestive effect on the choice of control–burner linkage arrangement by subjects, while the sign codes, ☆, #, △, and □ are immune to this suggestive effect.

2. The most frequently chosen arrangement by Chinese subjects was Type III, and no significant difference was found between Chinese male and female subjects.

3. If the suggestive effect is eliminated from the Shinar and Acton data, Type IV would become the most frequently chosen arrangement by American subjects.

4. The population stereotype of Chinese subjects in the control–burner linkage relationship was different from that of American subjects.

4.2. Physical model experiment

1. The different types of control–burner linkage affect the performance measures significantly.

2. Based upon the actual performance in reaction times and error rates for the experiment, Type V was the preferred arrangement for the Chinese subjects,

and Type II was the preferred arrangement for the American subjects.

3. The preferred arrangement of the physical model experiment was not in agreement with those of the computer simulation or paper–pencil test. This disagreement may be caused by the controls and burners not being in the same plane.

4. Therefore, using a computer simulation or paper–pencil test to study the compatibility or population stereotypes for the physical arrangement of displays and controls may not be effective.

Finally, the following recommendations are made:

1. As mentioned by Hsu and Peng (1993), for the sake of methodological soundness an experimenter should exclude the suggestive effect of the sequential codes in the study.

2. For the purpose of enhancing performance and reducing human error the difference of population stereotypes between Chinese subjects and American subjects must be taken into account.

3. It is probably best to use actual performance on a physical model to investigate the preferred arrangement between controls and displays.

Acknowledgements

This work was supported by the National Science Research Council, Taiwan, ROC, Project No. NSC 81-0115-E-131-14.

References

- Chapanis, A. and Lindenbaum, L.E., 1959. A reaction time study of four control–display linkages. *Human Factors*, 1(4): 1–7.
- Hsu, S.H. and Peng, Y., 1993. Control/display relationship of the four-burner stove: A reexamination. *Human Factors*, 35(4): 745–749.
- Osborne, D. and Ellingstad, V., 1987. Using sensor lines to show control–display linkages on a four burner stove. *Proc. Human Factors Society 31st Annual Meeting*, Santa Monica, CA. Human Factors Society, pp. 581–584.
- Ray, R.D. and Ray, W.D., 1979. An analysis of domestic cooker control design. *Ergonomics*, 22: 1243–1248.
- Sanders, M.S. and McCormick, E.J., 1992. *Human Factors in Engineering and Design*. McGraw-Hill, New York.
- Shinar, D. and Acton, M.B., 1978. Control–display relationships on the four-burner range: Population stereotypes versus standard. *Human Factors*, 20: 13–17.