SAFETY SCIENCE Monitor
Special Edition 1999
Safety in Action
25-28 February 1998
Philosophy of Safety
Article 6
VOL 3

A COMPARISON BETWEEN SAFETY INSTRUCTIONS AND PAYOFF MATRICES AT CHANGING TENDENCIES FOR USING SAFE PRACTICES

HEIDI-MARGUERITE BUSHELL & LENARD I. DALGLEISH
School of Psychology, University of Queensland
email: heidi@psy.uq.edu.au

ABSTRACT

A major focus of any safety program is the encouragement of the use of safe practices. Workers must make a decision about the acceptability of risk in any given situation. Where the employee makes the decision not to accept the level of risk, safe procedures may be used to reduce the risk. This decision process is influenced by two factors: 1) perceived probability of an incident occurring, and 2) the perceived costs and benefits associated the use of safe or unsafe behaviours. As the perceived probabilities of incidents is relatively low, workers may have a tendency toward not using safe practices, thus workers must be encouraged to use safe practices by other means. This study examines the use of safety instructions and payoff matrices of rewards and penalties encouraging safe behaviour. Results found that the frequency of safe practices in a simulated fire fighting task was increased more by payoff matrices which reward false alarms, than payoff matrices which penalise misses, or through the use of safety instructions.

Undoubtedly, the most effective way to create a safe working environment is to remove any hazards altogether. Unfortunately, not all hazards can be removed easily, so necessitating the use of safety programs, which reduce the risks to workers. Safety programs often include rules and/or training regarding the use of safe practices as part of the risk management process. Unfortunately, the implementation of rules and training may not sufficiently reduce unsafe practices, as the mere availability of safe procedures in the workplace is not enough to ensure safety (See Feeney, 1986).

Feeney's (1986) study found that workers avoid wearing Personal Protective Equipment (PPE) where poor functional design of equipment leads to discomfort or when they create interference with work tasks. Unfortunately it is the case that many safe practices slow down or inhibit the carrying out of production tasks. The use of safety rules does not necessarily allay this problem as safety rules are often seen to apply only in certain situations and as being impossible to follow in the many exceptional situations which are seen to be the reality of the shopfloor situation (Hale, 1990). Any time where production is perceived to compete with safety, the temptation is to neglect safety procedures so that work output can be increased (Quinlan & Bohle, 1991). Due to these competing priorities of safety and production (or comfort), employees may attempt to identify situations when safe practices are not needed. Thus, the decision to use safe practices some of the time, may be viewed as a task where workers attempt to discriminate between situations which require safe practices and those which do not. Viewing safety behaviour in this way, that is as a detection task, enables Signal Detection Theory (SDT) to be used to describe workers' performance.

A worker's response to a hazardous situation can be classified into one of four categories using SDT terminology. A situation where safe practices were used by the worker and an incident has occurred is known in SDT as a *hit*, where safe practices were not used and an incident occurred is known as a *miss*. Where safe practices were used but no incident occurred is referred to as a *false alarm* and where safe practices were not used and there was no incident, that is taking a short cut where it did not result in any negative outcome, is the *correct rejection* of safe practices. These outcomes can be summarised in a 2x2 matrix (see Figure 1). In this diagram, safe practices are defined as being 'needed' or not in accordance with the presence or absence of an incident.

<table>
<tr><td></td><td></td><td colspan="2">Actual State</td></tr>
<tr><td></td><td></td><td>INCIDENT
Safe Practices 'Needed'</td><td>NO INCIDENT
Safe Practices 'NOT Needed'</td></tr>
<tr><td rowspan="2">Worker's Response</td><td>Safe Practices Used</td><td>Hit
True Positive</td><td>False Alarm
False Positive</td></tr>
<tr><td>Safe Practices NOT Used</td><td>Miss
False Negative</td><td>Correct Rejection
True Negative</td></tr>
</table>

Figure 1 Matrix demonstrating actual alternatives and possible responses

Each of these outcomes has its own set of consequences. Even in the hazardous workplace, incidents are relatively rare, so that false alarms and correct rejections of safe practices will be the most common occurrence. There are many small but positive reinforcements for correct rejections of safe practices (e.g. taking shortcuts), and unfortunately there are many negative reinforcements for a false alarm (e.g. discomfort, extra effort required). The consequences of a miss could include a wide range of outcome potentials; from the most frequent superficial injuries, to the more rare occurrence of serious injuries and fatalities. This is referred to as the risk pyramid (Glendon & McKenna, 1995). In reducing the frequency of misses, the likelihood of common and costly superficial injuries, and of more rare and serious injuries are both reduced. Accidents are rare events that result from one of perhaps thousands of unsafe behaviours (Petersen, 1997). By reducing the frequency of unsafe behaviours, the antecedents of potential accidents may be reduced. The aim of this study is to compare two different means of attempting to change response tendency so that workers become more willing to use safe practices, 1) safety instructions and 2) payoff matrices which differentially reward and penalise safety behaviour.

According to SDT, response tendency can be affected by two sources: 1) the perception of the probability that a particular situation will be occur (i.e. p(incident)), and 2) by the costs and benefits incurred by a payoff matrix (Coombs, Dawes & Tversky, 1970). The effect of perceived base rate on response tendency, is through the individual setting the decision threshold so as to maximize the number of correct responses (McNicol, 1972). Consequently, if there are more situations requiring a particular response (i.e. not requiring safe practices) then the individual will become more willing to provide that particular response in the future (i.e. not using safe practices). Thus the relatively rare occurrence of incidents in the workplace may reduce workers' willingness to use safe practices. Consequently, workers' willingness to use safe practice may need to be increased by other means.

The costs and benefits for the consequences of the four different types of outcome (See Figure 1) also have an effect upon response tendency. In setting the decision threshold, human beings also try to maximize rewards and minimize penalties (McNicol, 1972). Different decision thresholds lead to different frequency of hits or false alarms. Consequently, if misses are sufficiently penalised or hits sufficiently rewarded, individuals will shift their decision threshold so as to increase the frequency of hits and false alarms. Individuals will attempt to maximize gains and minimize losses to compensate for the costs of misses and false alarms and the value of a hit or correct rejection (McNicol, 1972). Consequently, where a particular response is rewarded (i.e. safety), workers are more likely to respond accordingly (i.e. by using safe practices). The ‘reward’ may reflect "internal" values and ethics, or may be "external" such as monetary rewards or feedback. It must be remembered that the consequences of the four possible outcomes have values attached to them. While training or instructions in safety may make attempt to make the costs of certain outcomes (i.e. miss) more salient, safety programs which incorporate some form of behaviour modification in the form of a payoff matrix will provide more consistent reinforcement of these outcomes. Because of the direct effects of payoff matrices on response tendency, this study predicts that they will be more effective at increasing the tendency to use safe practices than instructions about the use of safe practices. Furthermore, it is predicted that instructions, and some payoff matrices will be less effective at changing response tendency under low base rate conditions than under high base rate conditions, due to the effects of low base rates on reducing willingness to use safe practices.

Because it is difficult to gain “gold standard” information about the actual occurrence of situations for which there were incidents and to the dubious ethical considerations of waiting for incidents to occur, it is difficult to test this in the field. However, a simulation can be constructed in the laboratory. A hazardous work situation where a choice may be made to use safe or unsafe practices may be simulated in a task where participants are assigned the responsibility of sending either 1 or 2 Fire Fighting Units (FFU) to fight fires in a computer task. The fire fighting task requires subjects to identify which of two possible alternatives is correct, that is subjects are asked to discriminate between 1FFU and 2FFU scenarios. The use of 2FFU to extinguish a fire may be viewed as a safe practice, as there is less probability of fire fighters being hurt due to lack of assistance or resources, however it is not always necessary to send 2FFU to a fire and it may be less costly to send only 1FFU. This is similar to the situation facing the worker in a hazardous workplace where safety is often in conflict with production.

Subjects’ responses on a trial will be classified as; hit, miss, false alarm, or correct rejection. Over a number of trials estimation of hit and false alarm proportions and from this response tendency will be calculated. Base rate will be manipulated through different probabilities of situations requiring 2FFU. Payoff matrices will be manipulated through the awarding of points. A payoff matrix will be used to induce a tendency toward sending 2FFU (i.e. safe practices) by incurring a larger penalty for a miss than for a false alarm, while both correct responses (i.e. hit and correct rejection) are rewarded equally (See Figure 2, Matrices B & C). A neutral payoff matrix will be created by equally penalising both incorrect responses (Figure 2, Matrix A). Half of the participants will only receive the neutral payoff matrix and will receive instructions about safe practices for the fire fighters under their command. The purpose of this is to examine how much safety instructions and payoff matrices shift the tendency to use safe practices.

METHOD

Participants. Participants were 24 students from the University of Queensland who were given credit points towards a first year Psychology subject for 5 hours of experimentation time, as well as the chance to compete for a $10 for the highest score each week and $50 for the highest overall score.

Materials. NEWFIRE, is written in Smalltalk/V286, a RAM resident, object-oriented DOS software system, which was installed on a 486 processor.

Design. The experiment involved three base rate groups (i.e. $p(\text{yes}) = .25$, $p(\text{yes}) = .50$, $p(\text{yes}) = .75$). Half the participants within each base rate were exposed to different payoff matrices without any safety instructions and the other half to safety instructions with a neutral payoff matrix. There were four participants in each condition.

Task. For both the task-training phase and the testing phase, participants sat in front of a computer screen showing a map of a forest area in the form of 18 x 18 green squares. When fire was reported in a given square, that square turned red. Participants used this information to base their decisions when sending out fire fighting units, which then reported back about their positions and activities. The testing phase utilised DUALFIRE which was a dual-choice fire fighting task, where participants were required to differentiate between scenarios requiring one fire fighting unit (FFU) and those requiring two units. The cues enabling this distinction were: 1) distance from the initial fire to the nearest available unit, 2) windspeed, and 3) meadow/no meadow (Loevberg, 1993). The further away the unit was from the fire, the greater the windspeed, and absence of meadow the more likely a fire required 2FFUs.

Base Rate. Each of the three groups were exposed to trials of different base rates. One group was exposed to trials where 25 out the 100 trials required a 2FFU response (i.e. p(2FFU) = .25). Another group was exposed to trials where 50 out of the 100 trials required a 2FFU response (i.e. p(2FFU) = .50). While a third group was exposed to trials where 75 out of the 100 trials required a 2FFU response (i.e. p(2FFU) = .75). Subjections were not informed of the base rates and were not given trial by trial feedback.

Payoff Matrices. At the beginning of each testing session participants were informed of the particular cost/benefit matrix for that block of trials. A neutral payoff matrix was used to establish a baseline, consisting of a loss of 10 points for an incorrect response (i.e. false alarm or miss) and gain of 10 points for each correct response (i.e. hit or correct rejection). Half of the participants in each group maintained the neutral payoff matrix throughout the experiment.

The remaining participants received three different forms of cost/benefit payoff matrix, to encourage a tendency toward responses of sending 2FFUs (see Figure 2). The first incurred a penalty of 20 points for each fire requiring 2FFUs where only one was sent (i.e. miss) (Matrix B) while the second incurred a penalty of 40 points (Matrix C). The third matrix encouraging a tendency of 2FFUs, provided a reward of 10 points for sending 2FFUs whether 1 or 2 were required (i.e. hits and false alarms) and reduced the reward for correct rejections (Matrix D).

Safety Instructions. The participants who received only neutral payoff matrices also received safety instructions. This group was instructed upon the components of risk, the cost of workplace accidents to Australia, and of the dangers to the individual fire fighters.

Measurements. Participants indicated choice of response of either 1 or 2 FFUs to the site of the fire and confidence ratings.

Instructions. Instructions were read to each participant describing the design and goal of the task. The appropriate cost/benefit payoff matrix was placed before the participant and explained to them.

Neutral

A.	2FFU needed	1FFU needed
2FFU sent	+10	-10
1FFU sent	-10	+10

Tendency to SAFE practice (Penalty for Miss)

B.	2FFU needed	1FFU needed
2FFU sent	+10	-1
1FFU sent	-20	+10

Tendency to SAFE practices (Harsh Penalty for Miss)

C.	2FFU needed	1FFU needed
2FFU sent	+10	-1
1FFU sent	-40	+10

Tendency to SAFE practices (Reward for False Alarms)

D.	2FFU needed	1FFU needed
2FFU sent	+10	+10
1FFU sent	-20	+5

Figure 2 Payoff matrices used to induce different response tendencies.

Procedure. The first 30 minutes consisted of training, using the general mode involving eight fire fighting units. In the testing sessions, participants were presented with five blocks of 100 trials each in the DUALFIRE mode, involving the choice of either 1 or 2 fire fighting units. Participants attended three sessions, each lasting 1 hour. The first session consisted of task training, and gaining baseline measures, in which the payoff matrix was neutral and the signal presentation probability was p(2FFU) = 0.5. The testing phase consisted of 5 blocks of 100 trials, each block taking approximately 30 minutes to complete. Payoff matrices were varied in the 5 blocks. These blocks were presented in a counter-balanced order as determined by a 5x5 latin square (Cochran & Cox, 1957). The participants were given feedback about the correctness of their decisions at the beginning of the subsequent session.

Results and Discussion

This study examined the effects of response tendency, manipulated by two different means; 1) payoff matrices and 2) safety instructions, upon the decision to use safe practices (i.e. send 2FFUs). Comparisons were made across three different base rate conditions. As the purposes of this discussion is to examine how much safety instructions and payoff matrices effect tendency to use safe practices, results from only some of the conditions are examined. The effect of the manipulations can be observed through the logarithm of the β value (ln(β)) indicating response tendency, and the proportion of trials participants used safe practices (p(yes)). Values of d′ were also included as a measure of sensitivity. A low d′ value indicates less ability to discriminate between situations requiring 1 or 2FFUs. Manipulations were not expected to change sensitivity.

The β values were calculated as a proportion of hits and false alarms, then a logarithmic transformation was performed resulting in values for ln(β). Values of ln(β) are symmetric around 0, so that negative values of ln(β) indicate a tendency toward safe practices while positive values of ln(β) indicate a tendency toward unsafe practices. For example, a value of -.2 would indicate a tendency toward safe practices while a value of .2 would indicate a tendency of the same magnitude toward unsafe practices. A proportion of greater than .50 indicates a greater use of safe practices (i.e. sending 2FFU).

The Effect of Safety Instructions on Tendency to Use Safe Practices

The means of the ln(β) values, the proportion of using safe practices, and the d′ values for participants given safety instructions can be seen in Table 1.

Table 1 The effects of safety instructions on mean d′, mean ln(β) values, and mean proportion of safe practices used.

Base rate ⇓		Baseline	Baserate	Baserate and Safety Instructions
p(2FFU) = .25	d′	1.09	.95	1.73
	ln(β)	.103*	.281*	-.492**
	p(yes)	.46*	.33*	.55**
p(2FFU) = .5	d′	.81	.89	.83
	ln(β)	.432*	.496*	-.592**
	p(yes)	.42*	.38*	.90**
p(2FFU) = .75	d′	1.63	1.44	1.60
	ln(β)	.095*	.282*	-.651**
	p(yes)	.44*	.56	.73**

* response tendency toward unsafe practices **response tendency toward safe practices

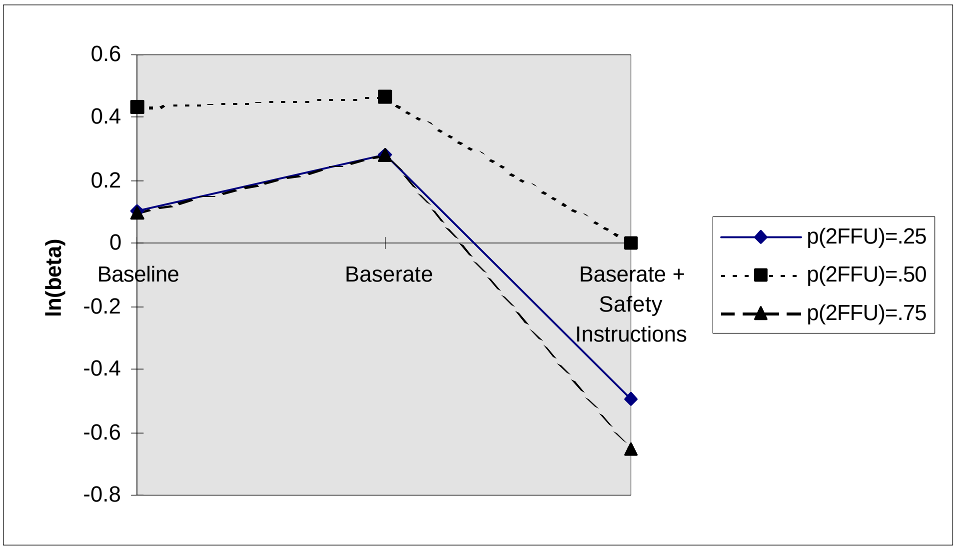

Figure 3 The effect of safety instructions on mean response tendency (ln(beta)).

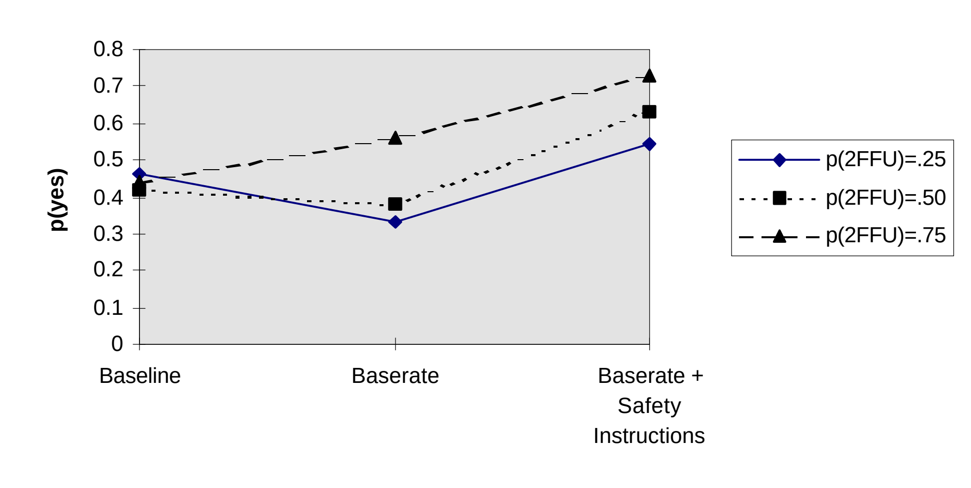

Figure 4 The effect of safety instructions on the mean proportion of safe practices used.

The response tendency for the Neutral Baserate condition was expected to display ln(β) values close to 0, as there was no response bias manipulation in this condition. However, it was expected that this neutral condition would reflect any response tendency inherent in individual participants, resulting from individual "internal" payoff matrices. From Figure 3 it can be seen that participants began with a slight tendency toward unsafe practices (i.e. points above the line). The manipulation of baserates (Grouped Baserate) with the same neutral payoff matrix, had inconsistent effects upon response tendency, with a slight tendency to further decreasing willingness to use safe practices, but the effect of base rate was not significant. As indicated by the negative ln(β) values, safety instructions successfully increased the willingness to use safe practices (i.e. points below the line). An analysis of variance revealed these differences in mean ln(β) values due to safety instructions to be significant ($F_{(2,18)} = 9.95$ $p < .01$).

These results are also reflected in the effect of safety instructions on the mean proportion of using safe practices (p(yes)). Figure 4 illustrates a significant increase in the proportion sending 2FFUs due to safety instructions ($F_{(2,18)} = 26.69$ $p < .001$). There was no significant differences in d′ for any of the conditions, indicating that sensitivity to the task did not change. From these results, it is evident that the use of safety instructions has shifted created a shift in tendencies from a lack of willingness to use safe practices to a willingness to use safe practices, without increasing ability to discriminate between situations which require safe practices from those which do not.

The Effect of Payoff Matrices on Tendency to Use Safe Practices

The following set of results report on the effectiveness of a means of creating a tendency toward the use of safe practices without any explicit instructions that this is the required response. The following conditions use payoff matrices which reward safe practices and penalise unsafe practices. The means of the ln(β) values, the proportion of safe practices used (p(yes)), and the d′ values for the different payoff matrix conditions are listed in Table 2.

Table 2 The effects of payoff matrices on mean d′, mean ln(β) values, and mean proportion of safe practices used.

		No Safety	Instructions		
Payoff matrix *Source of response tendency* **Base rate ⇓**	⇒	**A. Neutral** p(2FFU)=.5) for all groups *Neutral*	**B. Penalties for Unsafe Practices** *Safe*	**C. Harsh Penalties for Unsafe Practices** *Safe*	**D. Rewards for Safe Practices** *Safe*
p(2FFU)=.25	d′	.62	.55	.46	-
Unsafe	ln(β)	.161*	-.489**	-.452 **	-
	p(yes)	.42*	.79**	.82**	.97**
p(2FFU)=.50	d′	.69	.57	.70	-
Neutral	ln(β)	.371*	-.395**	-.176**	-
	p(yes)	.51	.79**	.66**	1.00**
p(2FFU)=.75	d′	.52	.46	1.33	-
Safe	ln(β)	.085*	-.488**	-1.237**	-
	p(yes)	.44*	.80**	.92**	.99**

* response tendency toward unsafe practices **response tendency toward safe practices

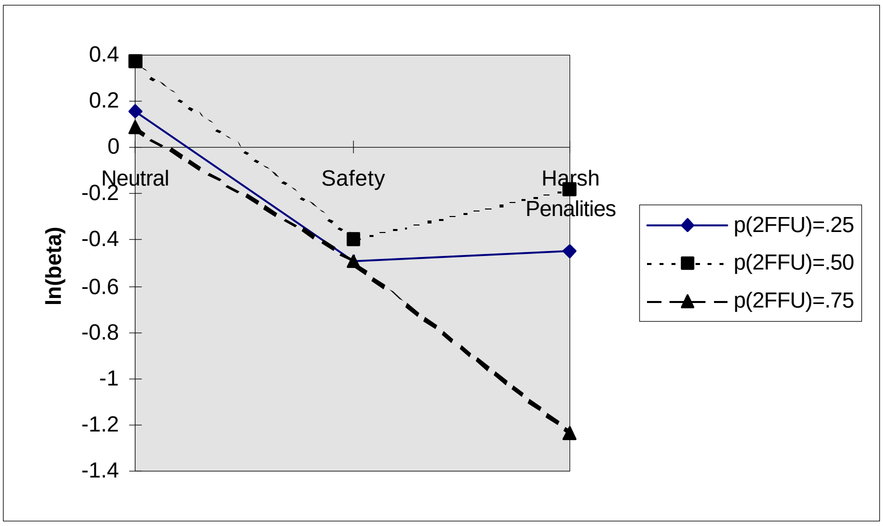

Figure 5 The effect of payoff matrices on mean response tendency (ln(beta))

Once again participants began the task with a slight bias toward unsafe practices as indicated by the ln(β) values above the line, but the payoff matrices penalising unsafe practices significantly reduced the mean ln(β) values to below zero ($F_{(2,16)}$ = 7.80 $p<.01$). Thus payoff matrices successfully changed willingness to use safe practices without the use of safety instructions.

Likewise, probabilities of using safe practices increased due to the use of payoff matrices (See Figure 6). The differences for the mean proportion of using safe practices for the different payoff matrix conditions was also found to be significant ($F_{(2,16)}$ = 34.96 p<.001). The results of the payoff matrix which rewarded safety, Matrix D, were not included in the analysis of variance, due to problems with the homogeneity of variance arising from all participants sending 2FFUs almost all of the time (p(yes)>.97). The very high p(yes) values also precluded calculations of d′ and ln(β) as there was no misses. The resulting ceiling effect, which is evident in Figure 6, while making calculations and analysis difficult is highly desirable in a safety program. Although, response tendencies, as indexed by the ln(β) values or p(yes), were expected to reflect the interaction of the two different sources; base rate and payoff matrix, there were no significant differences for baserate. Nevertheless, the ln(β) value where both the payoff matrix and the base rate were inducing a tendency toward safe practices (i.e. p(2FFU)=.75 combined with Harsh Penalties for unsafe practices, Matrix C) resulted in the next highest proportion of sending 2FFU (p(yes) = .92) after Matrix D (p(yes) = 1.00). The two sources of response tendency appear to have combined to result in a strong willingness to use safe practices. The different payoff matrices did not create changes in sensitivity to the task as indicated by the lack of significant difference for d′ values across conditions.

These results indicate that the use of payoff matrices successfully shifted tendencies towards a willingness to use safe practices. Because the means of doing this do not rely on internal and implicit means like the safety instructions, but provide external, explicit rewards and penalties the results are more consistent than those for safety instructions. As for safety instructions, the use of payoff matrices does not increase ability to discriminate between situations requiring safe and unsafe practices, but changes the threshold at which the individual is likely use safe practices.

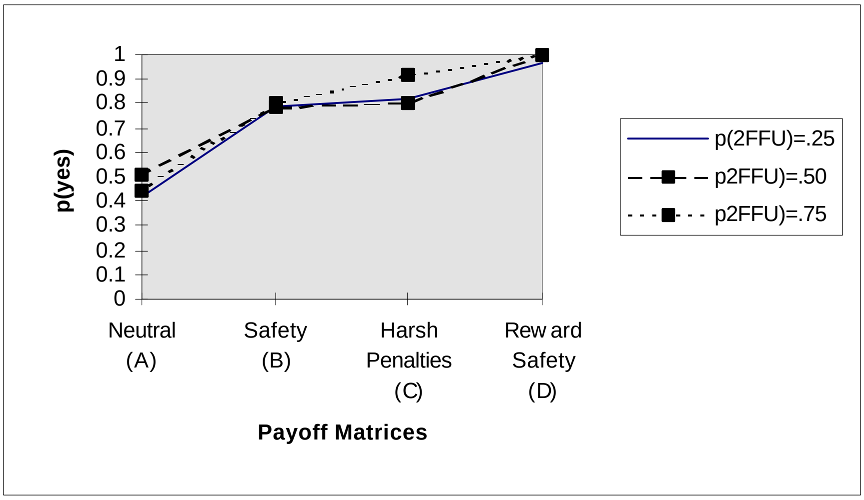

Figure 6 The effect of payoff matrices upon mean proportion safe practices used.

GENERAL DISCUSSION

This study has shown that payoff matrices produced a more consistent, and stronger willingness to use safe practices, than the use of safety instructions. While safety instructions may change “internal” values of the consequences of outcomes, payoff matrices provide a constant and explicit method of rewarding safe practices and penalising unsafe practices, so resulting in the desired change in behaviour. Although, one of the groups exposed to safety instructions also displayed a strong tendency to using safe practices (.90), the trend toward using safe behaviour was more consistent and successful when manipulated by ‘externally’ and explicitly by payoff matrices than ‘internally’ and implicitly by safety instructions.

One of the predictions for this study was that there would be significant differences according to perceived base rate. However, the effects of base rates on willingness to use safe practices should not be underestimated from the results of this study. The low base rate used here, p(incident) = .25 is a great deal higher than in a hazardous workplace, and workers receive a lot more feedback about the outcomes of their actions and have some perception of the probability of an incident in their work. The lack of effect for base rate may have been due to a lack of trial-by-trial feedback. Trial-by-trial feedback may have alerted participants to the fact that they were rarely being penalised for using unsafe practices (i.e. sending 1FFU) under the low base rate condition. Alternatively, perceptions of base rates may be altered through instructing participants about what kinds of base rate to expect. Past studies using this task seemed to indicate that participants were probability matching according to base rate. Future research will examine the effects of instructions which give actual and false expectations of the base rate, in interaction with payoff matrices. The payoff matrix rewarded for safe practices appears to have been more effective even those which penalised incidents where no safe practices were used. The effectiveness of rewards is consistent with the literature on behaviour modification which suggests that incentives (i.e. rewarding positive behaviour such as hits and false alarms), including monetary, praise and feedback, and team competitions improved safety and/or reduced accidents without exception, supporting the notion that rewarded behaviour tends to be repeated (McAfee & Winn, 1989).

However, caution must be taken in the introduction of rewards or incentives in safety programs. Peters (1991) concluded that while safety incentive programs are quite simple to implement, they are apt to be ineffective if not implemented correctly, resulting in substantial but often short-term increases in safe behaviour (McAfee & Winn, 1989). One problem that may occur is that rewarding safety with trinkets may trivialise the use of safe behaviour (Peters, 1991). It must be remembered that there are intrinsic rewards for employees in improving safety. A safety program which provides incentives or rewards must avoid discouraging employees from reporting near misses or minor injuries. Furthermore, some types of incentive programs such as the yearly picnic contingent upon no Lost Time Injuries, are not effective at reinforcing the use of safe behaviours as they do not differentiate between safe and unsafe behaviour so that both safe and unsafe behaviour are rewarded (Bradford & Ryan, 1997).

REFERENCES

Bradford, D.M., & Ryan, R.F. (1997). Behind the mirror of safety. *Professional Safety. Dec,* 34-36.

Cochran, W.G. & Cox, G.M. (1957). *Experimental Designs.* New York: Wiley.

Coombs, C.H., Dawes, R.M., & Tversky, A. (1970). *Mathematical Psychology: An Elementary Introduction.* New Jersey: Prentice-Hall.

Glendon, A.I., & McKenna, E.F. (1995). *Human Safety and Risk Management.* London: Chapman & Hall.

Feeney, R.J. (1986). Why is there resistance to wearing protective equipment at work? Possible strategies for overcoming this. *Journal of Occupational Accidents, 8,* 207-213.

Hale, A.R. (1990). Safety rules O.K.? Possibilities and limitations in behavioural safety strategies. *Journal of Occupational Accidents, 12,* 3-20.

Loevborg, L., (1993). Proposal for NEWFIRE dual-choice experiments. Draft.

McAfee, R.B., & Winn, A.R. (1989). The use of incentives/feedback to enhance work place safety: A critique of the literature. *Journal of Safety Research, 20,* 7-19.

McNicol, D. (1972). *A Primer of Signal Detection Theory.* London: George, Allen & Unwin.

Peters, R.H. (1991). Strategies for encouraging self-protective employee behavior. *Journal of Safety Research, 22,* 53-70.

Petersen, D. (1997). Behavior-based safety systems: A definition and criteria to assess. *Professional Safety. Jan,* 30-32.

Quinlan, M., & Bohle, P. (1991). *Managing Occupational Health and Safety in Australia: A Multidisciplinary Approach*. Melbourne: Macmillan.