Rewards and road user behaviour:

An investigation of the effects of reward programs on safety belt use

PROEFSCHRIFT

ter verkrijging van de graad van Doctor
aan de Universiteit Leiden,
op gezag van de Rector Magnificus Dr. W. A. Wagenaar,
hoogleraar in de faculteit der Sociale Wetenschappen,
volgens besluit van het College van Promoties
te verdedigen op 27 januari 1999
te klokke 14.15 uur

door

Marjan Paula Hagenzieker

geboren te 's-Gravenhage in 1962

Promotiecommissie:

Promotores	prof. dr. L. J. Th. van der Kamp	
	prof. dr. J. A. Rothengatter	Rijksuniversiteit Groningen
Referent	dr. W. H. Janssen	TNO-TM, Soesterberg
Overige leden	prof. dr. W. A. Wagenaar	
	prof. dr. J. A. Michon	
	prof. drs. J. Moraal	Technische universiteit Eindhoven
	prof. dr. A. R. Hale	Technische universiteit Delft
	drs. M. J. Koornstra	SWOV, Leidschendam

Contents

Chapter 1

Safe behaviour in traffic: Changing behaviour through rewards

Reward and punishment: these are basic manners to influence behaviour. To stimulate and control (safe) behaviour in traffic, punishment has been applied for many years in various forms while rewards have hardly been used at all. This despite the fact that research has shown that rewards certainly have an effect in this area also. This introductory chapter reviews some theoretical considerations and empirical findings of reward activities used to stimulate safe behaviour in traffic.

1.1 Introduction

Accidents are complex events often with multiple causes. Several studies indicate that the 'human factor' directly contributes to many – estimates vary from 50–80% – traffic accidents. But while often "the error is undoubtedly human, the remedy need not be", as McKenna (1982) puts it. It has generally been accepted in recent years that to prevent human error not primarily the individual but the (road) environment needs to be changed in order to induce 'automatic' and sustainable safe behaviour. However, the environment is still far from fail-safe (see, e.g., Cairney & Catchpole, 1991; Hale, 1991; Koornstra et al., 1992; Sabey & Taylor, 1980).

Road users deviate from the desired, normative behaviour due to unintended errors and intended, deliberate violations. It has been postulated that violations of traffic

rules contribute to the occurrence of accidents more than other types of failure such as errors or 'slips' and 'lapses'. However, attempts to empirically relate – using questionnaire studies – the concepts of violations and errors to accident liability have not been successful in identifying the exact proportion of accidents preceded by willful deviation from the normative behaviour, and hence due to violation of the traffic rules (Reason, 1979; Reason et al., 1991; Parker et al., 1992, 1995; Underwood et al., 1997; Rothengatter, 1997). Nonetheless, these studies have shown that at least some proportion of driving behaviour concerns deliberate deviations from desired behaviour. To reduce this type of deviant behaviour, behaviour has to be modified such that road users behave more 'normatively', for example through attitude change (influencing 'internal motivation' by means of, for example information and media campaigns) or by modifying external consequences (see, e.g., Rothengatter, 1997). Decreasing unintended deviations from the desired behaviour can probably be better achieved by adapting the road environment such that the (driving) task becomes 'easier', accompanied by road user education and driver training to induce robust 'automated' and 'safe' driving routines.

It appears not to be an easy task to modify road user behaviour. One of the problems is that errors or deliberate deviations often do not lead to negative consequences (accidents), and unsafe behaviour is often not corrected by feedback. To the contrary, deviations from safe behaviour may cost in terms of increased risk, but may be beneficial in more immediate terms; e.g., when in a hurry, speeding may help to arrive in time, or can induce feelings of pleasure or excitement (cf. Zuckerman, 1992; Jonah, 1997). Providing feedback and external consequences – such as reward and punishment – contingent upon specific behaviours may shift the balance of individual gains (and social costs).

The distinction between rewards and punishment has roots in conventional dichotomies between the carrot and the stick, penalties and incentives, and benefits and costs. Many different definitions exist for terms such as reward, punishment, incentive, and disincentive. A chosen definition usually originates from a specific theoretical background or approach. In the behaviourist tradition a dominant term is 'reinforcer' (or 'reinforcement'), which is related to the concepts of reward and punishment: these are only reinforcers if they increase the probability of that response (Skinner, 1953; Dickinson, 1989).

The terms incentive and disincentive refer to general motivating conditions, while the terms reward and punishment refer to the consequences contingent upon certain behaviours (cf. Van Olst & Bakker, 1979). *Reward*, then, is a positive consequence contingent upon behaviour and *punishment* is a negative consequence contingent upon behaviour. A possible reward delivered as a consequence of behaviour X is an *incentive* to display behaviour X, but a *disincentive* to behaviour Y (and punishment is a disincentive for X and an incentive for Y, where X and Y are

mutually exclusive – e.g., overtake or not). A reward or punishment is not always a reinforcer, and a similar reward may have an effect on the behaviour of some people but not on others. For example, relatively small payments for 'good' or 'safe' road user behaviour (such as wearing a safety belt) may not be seen as a reward, while for instance additional unpaid vacation time for 'good' behaviour may actually be perceived as a punishment (cf. Winett et al., 1989). Rewards and punishments are only *reinforcers* if they do have an effect on the frequency of contingent behaviour.

It has been traditional practice in the traffic system for desired behaviour to be stimulated by rules, regulations, and laws, resulting in the punishment of offences. Many studies haven been conducted to investigate how penalties and enforcement can effectively modify individual behaviour in traffic (see, e.g., Goldenbeld, 1993; Grant et al., 1991; Zaal, 1994). Nevertheless, there are still many gaps in the knowledge about the most effective enforcement methods. On the other hand, good or appropriate behaviour is rarely highlighted in road safety (Fildes, 1995, see also Friedland, 1989). Modifying road user behaviour by reward has not been given much attention in the practice of road safety, despite the fact that (psychological) theories offer some fruitful ideas about influencing road user behaviour by reward. Moreover, rewards have been widely used in other fields of application (e.g., industrial safety, health-related behaviour).

Relatively few (mainly small-scale) studies on the effects of rewards on road user behaviour are available, and even less is known about which methods are most effective. Therefore, modifying individual road user behaviour by reward is the topic of the present study. The main questions that will be addressed are whether rewards are effective in modifying individual road user behaviour, which types of reward (programs) are more effective than others, and which other conditions or circumstances mediate this effectiveness. The question whether reward programs can be successfully applied on a larger scale will also be addressed.

In the present study incentives will be viewed as limited to promised rewards intended to increase the occurrence of specific (road user) behaviours – and sometimes to attain certain 'outcomes', such as accident-free driving. (Note that rewards offered for achieving a certain outcome, for example accident-free driving, or for achieving a collective outcome, for example achieving accident-reduction in a given area or group of drivers, are not rewards in the strict sense, i.e. contingent upon a specific behaviour.) Defined in this way 'incentives' and 'rewards' are different terms referring to the same concept (cf. Dwyer & Raftery, 1991).* Various types of rewards can be distinguished. We will focus on external rewards that represent some material or financial value. Examples of this type of reward are

*Where the term 'incentive' is used in later sections and chapters it is used as equivalent to 'reward'.

exchangeable tokens, immediate valuables, promotional items, a chance to win a contest, and work-related privilege (see Geller, 1982). Non-material rewards, such as for example praise or social attention, and verbal feedback and prompting techniques will not be considered.

The first part of this introductory chapter sketches the role of reward in various theoretical approaches and in some applications to modify behaviour in other fields than road safety. In the second part of the chapter the role of reward in models of road user behaviour is addressed, and an overview is presented of the available research results into the effect of rewards on (safe) behaviour in traffic. The chapter concludes with formulating the research questions of this thesis, and provides an overview of the following chapters.

1.2 The role of reward in behaviour change: Some theoretical considerations

A multitude of psychological theories – generally in the field of learning and motivation – reserve an important place for the principle of reward. In theories varying from those postulated by Skinner's operant conditioning and Bandura's social learning theory to theories with a more cognitive orientation, such as Festinger's cognitive dissonance theory, subjective expected utility theory as well as attitude theory, rewards are explicitly regarded as powerful influences on behaviour, even though the principles and presumed mechanisms of effect of the individual theories can vary greatly (see for an overview, e.g., Lindzey & Aronson, 1985). No matter how different these theories may be, the effect of reward is generally considered substantial. The role of reward in these various approaches is briefly sketched below.

1.2.1 Operant conditioning

Rewards are most prominent in operant learning theory (Skinner, 1938, 1953). In operant, or instrumental conditioning, behaviour is modified by its consequences. In this type of learning the individual 'operates' on the environment. If this behaviour has a favourable outcome (is followed by a reward), there is greater chance the action will be repeated (the behaviour has been reinforced).

Operant behaviours generally occur only under specific conditions. Most behaviours are regulated by discriminative stimuli, environmental events which had been associated with reinforcers in the past. These stimuli signal the time or place

when an operant is likely to have favourable consequences. The police officer on the street is a discriminative stimulus indicating that actions complying with the law may produce favourable outcomes, whereas noncompliance is likely to have aversive consequences (see Berkowitz, 1980).

Behaviourists stress that a reward is not the same as reinforcement. Sometimes rewards do not lead to an increased frequency of the behaviour concerned. In this context, for example, the distinction made by Williams (1980) between attractive and unattractive rewards, has been criticized: "[Unattractive] rewards were not reinforcers in the first place because they did not enhance the target behaviour. It seems that highly reinforcing rewards (attractive ones) may result in post-reward performance increases, while less reinforcing or nonreinforcing rewards (unattractive) may result in post-reward decreases" (Dickinson, 1989).

Operant learning theory states that rewards (and punishments) should be contingent upon specific behaviour. When, for instance, the use of a safety belt has been rewarded in the past (or the non-use punished) this will not necessarily lead to a more frequent occurrence of other (desired) road user behaviours, such as stopping for a red traffic light, or keeping the speed limit. Furthermore, variables such as the frequency, probability, and timing of 'rewarding' (reinforcing) have an effect upon the occurrence of subsequent behaviour. For example, an immediate reward will increase the chance of a particular behavioural response much more than delayed, postponed delivery of the reward. The principle of partial reinforcement refers to rewarding the desired response only a small percentage of the time. Research has shown that such partial reinforcement will make a desired response much stronger and more resistant to extinction than will continuous reinforcement every time the response occurs (Skinner, 1953). Other variables, such as the size and type of the reward are also of importance. Tangible and large rewards are viewed as better reinforcers than small, intangible rewards.

1.2.2 Social learning

In Bandura's (1977, 1986) theory the social context in which behaviour is acquired and maintained is taken into account. Though Bandura's theory of social learning puts less emphasis on reinforcement and more on vicarious observational learning (by imitation without the necessity of reinforcement), it still maintains an important place in the theory for reinforcement as a facilitator of performance. In general, rewards are considered more effective in regulating behaviour that has already been learnt, than in creating completely new behaviours. Much of human behaviour is maintained by anticipated rewards rather than by immediate reinforcement. Cognitive processes interact with external factors such as reward and

punishment. By noting the rewarding or punishing consequences of their own actions or the actions of others, people develop cognitive expectations about behavioural outcomes and about what they must do to achieve desirable outcomes or to avoid unpleasant ones. So, in addition to the important role of reinforcement, behaviour patterns can also be learned through observation and imitation of others (social modelling). In the course of such observation, one may see others receive rewards for certain behaviours and consequently may later tend to behave in similar ways oneself.

External rewards and punishments are not the only source of reinforcement, according to Bandura. Many actions are regulated by self-imposed consequences. People develop their own values about what activities are important to perform and personal standards of behaviour (cf. Hall & Lindzey, 1978). These internalized standards lead them to evaluate their own actions and to reward and punish themselves by self-approval or self-criticism. What people have to learn is not simply that rewards are forthcoming, but that they can reach their goals through their own actions ('self-efficacy'). What individuals come to reward and punish in themselves may reflect the reactions that their behaviour has elicited from (important) others (e.g., parents, peers, and other socializing agents). It therefore matters, for example, who is delivering a reward (cf. Berkowitz, 1980).

Because the social environment is considered important in Bandura's theory, social pressure can act as an extra stimulation for the individual, for example in encouraging others to wear a safety belt in order to raise the achievement of a group. If rewards are dependent on the goal to be achieved by a group, in comparison with individual goals, their effect could be even greater. However, when competition is involved, this may sometimes be counterproductive (Berkowitz, 1980).

1.2.3 Social cognitive theories

Many theories can be considered to belong to the cognitive approach. To illustrate the role of rewards, we will focus here on three examples of approaches: the theory of cognitive dissonance, attitude theories, and subjective expected utility theories.

Cognitive dissonance theory states that cognitive dissonance is an unpleasant state of tension generated when a person has two or more 'cognitions' (bits of information) that are inconsistent or do not 'fit together' (Festinger, 1957, 1964; see for related theories, e.g., Deci & Ryan, 1985; Lepper & Greene, 1978). The theory holds that people will try to reduce dissonance by changing one or more of the inconsistent cognitions, by looking for additional evidence to bolster one side or the other, or by derogating the source of one of the cognitions. The greater the

dissonance, the stronger the attempts to reduce it (Oskamp, 1984). An external reward can reduce the cognitive dissonance that results when an individual engages in behaviour one is not (internally) motivated to, for example using a safety belt when one is not convinced of the usefulness of the safety belt. From cognitive dissonance theory follows that the reward should be large enough to induce behaviour change, but not so large that it can be reasoned that it was the sole motivator; as the behaviour then will extinguish when the reward is removed. The explanation given by Lepper et al. (1973) for the possible undermining effect of rewards is called the 'overjustification' hypothesis: If an external reward is offered and provided for engaging in an initially enjoyable task, the target activity is perceived by the person as overjustified because a reward is not necessary, and the individual infers that engaging in the activity was basically motivated by the external contingencies of the situation, rather than by any intrinsic interest in the activity itself (Lepper et al., 1973). On the other hand, one may initially engage in a behaviour because of an external reward, but later on quite willingly take on, for example, an onerous administrative task – not because one sees it as interesting or fun, but because one believes it is valuable and personally important for some reason (Ryan & Deci, 1996). In the area of road safety, one could similarly argue that, for example, certain people initially start to use their safety belts (mainly) because of the promise of an external reward, but maintain using their belts – even when rewards are withdrawn – because they view safety belt use as important for their personal safety.

Other useful insights are provided by *attitude theories*, and by research on the relation between behaviour and attitudes. While in many attitude theories it is assumed that attitudes influence behaviour (see, e.g., Ajzen, 1985), more recent theories also stress that the impact of behaviour on attitude formation can be considerable (see, e.g., Eagly & Chaiken, 1993; Olson & Zanna, 1993). In this view attitudes are the cognitive evaluations of outcomes of behaviour. Research has shown that contemplation of an attitudinal position can change during or in anticipation of engaging in a certain behaviour, particularly when people are exposed to positive incentives for adopting certain attitudes or when they are unconvinced or doubting about their current attitudes towards that behaviour. So, while rewards can influence one's attitude towards a certain behaviour and as a result also influence behaviour, rewards can also induce behaviour directly, resulting in a change towards more positive cognitions (attitudes) towards the rewarded behaviour.

Useful perspectives are also provided by *subjective expected utility* theories. Although these theories are usually not classified as (social) cognitive theories by representatives from this approach, it was decided to treat them in this section, because the choice of an individual is generally seen as the result of what can be called cognitive evaluations of alternative behaviours. These approaches describe the relation between the probability and the value of outcomes and the choices of

an individual (see, e.g., Lonero et al., 1994). Utility theories, decision-theory, cost-benefit theories (e.g., Kahneman & Tversky, 1979) and expectancy-value theories (e.g. Atkinson, 1964) all account for a person's choices in terms of the net expected utilities of various alternatives (see also Abelson & Levi, 1985; Heckhausen, 1991; Vlek & Wagenaar, 1979).

The basic assumptions of expectancy-value theory are in accord with common-sense thinking about motivational behaviour: What behaviour is undertaken depends on the perceived likelihood that the behaviour will lead to the goal and the subjective value of that goal (Weiner, 1992). Hence the greater the belief that the goal will be attained and the higher the incentive value of that goal, the greater the motivation tendency to engage in the appropriate instrumental behaviour. The assumptions made by expectancy-value theories are similar to those imposed by decision theorists in their analysis of action. In expected-value theories the expectancy and payoff values for each alternative are combined in a multiplicative manner and a decision is made based on subjective expected utility (SEU), or the expected personal value associated with each choice. Rewards and punishments can influence which action is taken, because these change the (perceived) utilities of actions by making the subjective benefits of safe behaviour greater than the costs.

In terms of strategies per sé, the evidence suggests that decision makers adopt different heuristics on different occasions, depending on the number and complexity of the available alternatives (Payne, 1976; cited in Vlek & Wagenaar, 1979). In a two-alternative situation, for example, individuals may consider each alternative in turn and arrive at a decision reflecting a trade-off among several dimensions. When a choice among alternatives is complex, however, persons often employ 'satisficing' (Simon, 1957, cited in Vlek & Wagenaar, 1979) rather than optimizing decision strategies. It also appears that people tend to overweigh certain and short-term advantages over uncertain and long-term risks (Kahneman & Tversky, 1984).

According to prospect theory (Kahneman & Tversky, 1979), people are generally more motivated to avoid risks than to obtain profits. A punishment of $100 for current behaviour or a reward of the same amount for alternative behaviour are economically equally strong 'incentives'. However, from the perspective of the individual the punishment of $100 is seen as a risk, whereas the reward of the same amount as possible gain. Schmidt (1989) states that prospect theory is presumably particularly relevant for breaking habits, while learning theory (in which reward is considered to be more effective than punishment) applies more to behaviours that are not yet habitual.

1.2.4 Conclusion

All theories treated in this section have in common that 'reward' is an important – sometimes the main – motivational construct in controlling and explaining behaviour. From the various approaches, variables such as the size and type of the reward, the probability and timing of rewarding, the type of task or behaviour one is rewarded for, the social setting (which agents are involved, who delivers the reward, the role of groups), cost-benefit considerations (in relation to the goal to be achieved), and other cognitive evaluations of these variables can be identified as important factors mediating the effects of rewards on behaviour. While in learning theories rewards are generally ascribed a main role in changing behaviour, social cognitive theories stress the subjective evaluations of outcomes of behaviour and – while acknowledging the potential of rewards to induce motivation to exhibit certain behaviours – also express reservations toward (large) extrinsic rewards. More generally, all social cognitive theories emphasize beliefs, perceptions, expectations, and information processing abilities as cognitive mediators of external contingencies. Characteristic in cognitive dissonance and related theories is the proposal that the undermining effect results from complex cognitive inferences involving an evaluation of internal and external reasons for engaging in the target activity.[†]

Research has shown that rewards may enhance motivation and performance when these rewards are *non-tangible* (such as verbal praise of task performance; e.g., Deci, 1975), when they are perceived as *attractive* by the rewardee (Williams, 1980), when task performance is not tied (*noncontingent*) to the reward, and when rewards are *self-administered* ('choose your own reward'; cf. Deci & Porac, 1978). In addition, when the task involves *less initial interest*, or *routine well-learned activities* (see, e.g., Sarafino, 1984), when there is *resistance to behavioural change* (McCullers, 1978), and when the task is *aversive* (McGraw, 1978) rewards can enhance performance. Extrinsic rewards could awaken people's intrinsic motivation for an activity if they lead people to activities they had never tried or if they help people to develop a level of competence necessary to enjoy the activity (Lepper et al., 1973). On the other hand, intrinsic motivation is more likely to be undermined by *salient* rewards (when the subject's attention is focused on it or the reward is clearly visible; see e.g., Sarafino, 1984), by rewards that are more *concrete or tangible*, such as candy or money (Deci, 1975), by rewards that are perceived as *unattractive* to the rewardee (Williams, 1980), by *expected* and *contingent* rewards (Lepper et al., 1973). In addition, extrinsic rewards tend

[†]In particular, the possible detrimental effects of rewards on behaviour and intrinsic motivation as an intervening variable of behaviour (more specifically: creativity and task interest) have – up to the present – been the topic of heavy debate (see, e.g., Cameron & Pierce, 1996; Eisenberger & Cameron, 1996; Kohn, 1996; Lepper et al., 1996; Ryan & Deci, 1996; for recent discussions on the topic).

to impair people's performance on *open-ended activities* such as problem solving (cf. McGraw, 1978; Condry & Chambers, 1978), in *competitions*, and when an (attractive) task requires *creativity* (e.g., Deci & Ryan, 1980).

In conclusion, while learning theories mainly emphasize the positive effects of tangible rewards, and utility oriented theories acknowledge the role of such rewards in modifying subjective expected utilities and thereby behavioural choices, (other) social cognitive theories also point at the possible negative effects of tangible rewards because those behaviours that are initially internally motivated could become mainly externally motivated as a result of external rewards. The issue appears to be not whether rewards should be used, but how and for what purpose, how they can be used most effectively and how their effectiveness should be evaluated. Engaging in what is perceived to be an interesting task is considered to be intrinsically rewarding; external rewards are not necessary and can be detrimental. To get a person do an uninteresting task, external rewards will surely help. External rewards will not always work, as people sometimes react against constraints or opt for paths that provide more freedom (Brehm, 1966). Still, rewards do often serve to control behaviour.

1.3 Rewards to modify behaviour: Some applications

In various fields of application rewards have been used to modify behaviour, for example in behaviour modification, the behavioural engineering approach, in the area of industrial safety, and in health settings. These will be reviewed in short.

1.3.1 Behaviour modification

Behaviour modification, applying concepts from learning theory, has typically been successfully applied in clinical (therapy) and school settings. It involves rewarding people when they behave in some desired manner and withholding rewards when undesired behaviour is exhibited. So-called 'token economies' are an example of the behaviour modification approach. In token economies appropriate behaviour is rewarded with tokens that can be exchanged later for the desired goods or activities. Measurement is crucial in behaviour modification and therapists or teachers follow a regular schedule: identify and specify the desired behaviour (not just 'being good' for example); count and record the frequency of the desired behaviour; reward the behaviour immediately when it occurs; and evaluate the results (Kazdin, 1984).

What is common to the many different formulations and techniques of behaviour modification is direct intervention to alter a person's reactions to situations that this person or some significant others deem changeworthy. Behaviour modification procedures are direct: disturbing and changeworthy behaviour is not presumed to be indicative of an 'underlying' disorder (Ullmann, 1984).

When rewards or other extrinsic controls are relied on to produce behaviour change, there is a high probability that the behaviour will become dependent on those rewards and will not persist in the absence of rewards. Behaviour modification can be very effective with people who have little intrinsic motivation when treatment begins. It has therefore been recommended that rewards be kept at the minimum level that will ensure the needed responding; whenever possible the target person should participate in establishing the contingencies, and when rewards are removed they should be withdrawn gradually so the people can develop internal controls to replace the external ones being removed (Deci, 1978).

1.3.2 Behavioural engineering

Behavioural engineering advocates an approach referred to as the antecedent-behaviour-consequence (ABC) model, whereby conditions preceding target behaviours (antecedents) and conditions immediately following target behaviours (consequences) are manipulated in order to motivate behaviour change (Geller, 1989a). The behavioural engineering approach has successfully been applied in stimulating environment-relevant behaviours, such as energy conservation and littering (e.g., Geller et al., 1982c, 1990; Midden et al., 1980).

It appears that antecedent conditions using commitment, demonstration, and goal-setting strategies were generally most effective in encouraging environmentally responsible behaviour, and consequence conditions (feedback, rewards and penalties) were effective in producing behaviour change during the experiment's duration (Geller, 1990). The consequences of positive reinforcement that have been applied to benefit environmental protection have varied considerably. Some consequences have been contingent upon the occurrence of a desired behaviour, whereas other strategies did not specify a desired response but were contingent on a certain outcome (e.g., on a certain obtained level of environmental cleanliness, energy consumption, or water savings). Both behaviour- and outcome-oriented strategies have shown to be effective in the area of environment-protective behaviour. However, most behaviour-change interventions were characterized as small-scale and lacking durability. Intervention effects in communities were often small compared to individual and small-group effects, and the effects were often transient (Geller, 1990; see also Dwyer et al., 1993). Some exceptions are

noted, usually because active local groups and volunteers were available to implement and help institutionalize the behaviour-change program; these concerned applications of this approach to the promotion of safety belts (Geller, 1989a). In addition, much of the research in this field did not directly compare interventions, and many potentially effective intervention strategies have not been applied (i.e., group interventions and penalties).

1.3.3 Industrial safety

Many companies operate safety incentive schemes of one sort or another, and many different forms of rewards and feedback have been studied. These variables can be categorized in terms of monetary incentives, praise and feedback, and team competitions which may also have involved the use of cash awards. Most studies have used these variables in combination, making it difficult, if not impossible, to isolate the impact of any given intervention variable on the dependent measures (McAfee & Winn, 1989). Most were based on the behaviour modification principle that rewarded behaviour is likely to be repeated. A few studies were based at least in part on goal-setting theory. Also a wide variety of dependent measures were used (e.g., Reber et al., 1984; Chhokar & Wallin, 1984; cited in McAfee & Winn, 1989). Many of these could be considered 'causal' variables (specific behaviours) whereas others were outcome variables (injuries/accidents). The major finding was that all studies found that incentives or feedback were successful in improving safety conditions or reducing accidents. In addition, reward programs have generally appeared to be cost-effective. Companies which have introduced reward programs have reported a doubling in cost savings on injuries and accidents in relation to the cost of running the program (Wilde, 1985, 1994), although it has also been argued that, if the behaviour requires considerable effort, if employees are only mildly concerned about the aversive consequences, or if the unsafe work habits are well established, it may be necessary to use more costly rewards to motivate the desired behaviour (Peters, 1991). The costs involved in this strategy include the cost of the rewards or privileges being offered as incentives, the time required to take periodic performance measurements on the target behaviours, and the time required to provide feedback. An unanswered question concerns the extent to which reward programs encourage employees to fail to report relatively minor types of accidents and injuries in order to avoid losing all or a portion of the reward being offered (Peters, 1991).

The results of these intervention studies, however, have often not been reported in the scientific literature but appear in the trade and professional press. So, the claims are usually not backed up by detailed analysis or argument (see Hale & Glendon, 1987). This is a great pity, according to Hale and Glendon, because

the overall results reported from the major schemes are impressive. For instance, 81% reduction in works injury accidents and 50% reduction in driving accidents over ten years have been reported (Bodycomb, 1986; cited in Hale & Glendon, 1987). Here a scheme was used with awards in the form of points exchangeable for goods for keeping below monthly and quarterly targets of accidents, and for specific achievements – for example, passing safety-related tests. A reduction in disabling injuries over ten years from 114 to 3 per year (Fox et al., 1987) was found in a similar scheme in which vouchers for goods were awarded for meeting specific targets related to time free of disabling injury and other specific criteria. The authors usually do not claim that all of the effect comes directly from the incentive scheme, since many design and work procedure changes have usually taken place over the same period. Hence it is very difficult to disentangle the incentive effects from those of other activities. The reported improvements, however, are far too impressive to dismiss lightly.

In general, it appears that the easier it is to reliably measure performance of the (self-protective) behaviour, the more effective is the strategy of incentives and disciplinary actions. It is also noteworthy that the interventions which are reported as successful are those that have specific and attainable targets linked with individual and group behaviour, with carefully thought out and significant prizes which keep interest alive, and always give a target to aim for even if a first one is missed. The successful schemes all involve elaborate targetsetting and monitoring (Hale & Glendon, 1987; McAfee & Winn, 1989; Peters, 1991). It has also been noted that long-term effects have hardly been studied and that not all parts of the programs were equally successful. Peters (1991) also concludes that some industrial incentive campaigns appear to fail. It also appears that it is far from clear which incentives are more effective than others; how environmental, individual and task characteristics influence the impact of incentive schemes, and how long incentives will continue to motivate safety behaviours when the program is run on a continuous basis.

1.3.4 Health-related behaviour

Many recent health promotion programs are based on goal-setting approaches; some of those also involve rewards. The question can be raised whether external rewards enhance goal attainment (Strecher et al., 1995). It appears that rewards have been effective in stimulating certain preventive health precautionary measures, such as weight control and smoking cessation programs (see, e.g., Burke et al., 1987; Jeffery et al., 1993; Stevens et al., 1996; Warner & Murt, 1984). These are usually work-site or school-based health promotion programs. A variety of rewards have successfully been applied, such as extra payment, time-off, and both individual and group-oriented reward programs have been used.

However, a number of potential long-term problems may occur when building an incentive system into a goal-setting program. First, the reward system in a clinical setting will probably not exist in the real world. Transfer of reward systems into normal environments is often discussed but rarely accomplished. Second, providing incentives may have a tendency to externalize the source of motivation, ultimately reducing goal commitment and performance. If provided in such a way that commitment is not undermined, however, incentives can play a role in motivating persons to set goals (Strecher et al., 1995).

1.3.5 Conclusion

Based on the described applications of rewards in various settings, a number of conclusions and recommendations for effective use can be formulated (see also Balch, 1980; Hale, 1991; Winett et al., 1989). Reward programs have generally been successfully applied in small-scale settings (clinical, school, industrial, worksite applications, and in local communities). Larger-scale applications have rarely been tried, and those that have been studied appear to have been less effective. Reward programs appear to be especially effective in reaching substantial short-term effects. Both behaviour and outcome-oriented reward programs appear to be effective. Simple, direct pay-off contingencies appear to be most effective, i.e. what behaviour or outcome leads to reward. The target group should have (or be given) the necessary knowledge and skill to exploit the contingencies, and should believe that their attainment is dependent on their own behaviour. This seems to be particularly relevant for outcome-oriented reward programs. Concrete desired outcomes, i.e. specific and detailed targets (preferably in consultation with the participants) and rapid and clear feedback about progress to the goals, appear to enhance the effects of the program. It is therefore necessary to systematically measure and monitor the target behaviour, to be able to evaluate (and when necessary modify) the reward program.

Both individual and group-oriented programs appear to be successful. Some have stressed that group targets enhance social motivation, and could therefore lead to better results (Hale, 1991). In addition, it appears to be important that the rewards themselves are perceived as meaningful and attractive to the participants, and that the magnitude of the reward is sufficient to promote the behaviour change. However, overly large incentives may create a situation where subsequent, more appropriate incentives will be perceived as too small. They should be seen as equitable by the participants, and care should be taken to maximize the delivery of incentives so that receipt of incentives does not appear subject to chance (i.e. as with some lotteries). Immediate delivery of rewards is generally seen as important, especially during the initial stage of the program. Furthermore, rewards should be

progressive; one failure should not lead to all subsequent rewards being unattainable, but continued success should be more highly rewarded as it nears the target (on the assumption that more effort is needed to make improvements as the ceiling is approached). Some have stated that rewards should most often be used for the acquisition stage (Winett et al., 1989). The program as a whole should be launched and run with a high profile to arouse interest and make the target behaviour as well as the reward activities salient (optimizing the vicarious learning by seeing others being rewarded). Finally, rewards are generally seen as only one part of an overall program and comprehensive approach.

If one considers to apply reward programs in the area of road safety a number of problem factors can be anticipated. These concern the type of behaviour to be rewarded, the scale of the reward program, the necessity to systematically and carefully monitor the behaviour concerned, and the organisation and agents that set up the program and deliver the rewards, respectively.

Apart from effectiveness, what can be called 'ethical' considerations of applying reward programs can play a part; these are usually considered less relevant when applying, for example, punishment and motivational approaches not involving rewards. The question can be raised whether behaviour that is legally obligatory should or must be rewarded. For example, in many countries it has been obligatory to wear seat belts in a car, and laws specify that car drivers are not allowed to have a blood alcohol content of more than a given promillage (usually 0.5 or 0.8). As far as this is concerned, there are no objections to stimulating desirable, though voluntary, behaviour by rewarding such behaviour. Can it, however, be imagined that government bodies actually reward such desired behaviour? They are in fact the same bodies that are supposed to punish breaking the law. Moreover, 'punishment' (police enforcement) is built-in into the traffic system, whereas rewards are not. Reward programs require careful monitoring of the behaviour concerned. How could reward programs be implemented in the traffic system, and which organisation or agents should be responsible for the delivery of rewards? One possibility could be that private companies or organisations are used to encourage desirable road user behaviour by offering rewards. In addition to whether reward systems can be efficiently implemented in the traffic system to enable large-scale applications, the question must be raised whether such large-scale applications can be effective in substantially increasing desired road user behaviours. As stated earlier, reward programs have been particularly successful in small-scale settings.

On the other hand, a number of advantages of reward programs appear to exist in comparison to other approaches not involving rewards. For example, whereas it has been shown that information and media campaigns often do not lead to behaviour change (see, e.g., McGuire, 1985), reward programs have shown to be effective in accomplishing substantial behaviour change, in particular in the short

term. Furthermore, reward programs may be less costly than enforcement programs, requiring comparatively less manpower and administrative actions when applied in small-scale environments. In industry, reward programs generally have appeared to be cost-effective. If workers demonstrate safer behaviour on the road, for example, this could reduce absenteeism and (indirectly) increase productivity, as fewer injuries and accidents occur (Geller, 1984b).

Another advantage of rewards as compared to 'punishment' is that reward programs are, in general, considered as measures that are rated favourably among the public, as questionnaire studies have shown. For example, a Canadian study, in which road users were requested to classify some 40 road safety measures according to popularity or desirability, showed that incentive campaigns were considered one of the most favoured measures (Wilde et al., 1975; see also Harano & Hubert, 1974). Other studies, conducted in various European and North-American countries, have also shown positive attitudes towards reward programs (e.g., Caverson et al., 1990; Machemer et al., 1995; Mortimer et al., 1990; Sagberg, 1994). In general, a (large) majority of the respondents in these studies were in favour of rewards to stimulate safe road user behaviour. Rothengatter et al. (1992), in a report on the social acceptance of automatic policing and information systems, postulate that acceptance of incentive-based systems by road users is very likely to be much higher than fine-based systems, as they bypass most of the negative consequences attached to fine-based systems.

Having reviewed the role of reward in some theories and applications outside the realm of road safety, pointing at possible advantages and disadvantages of reward programs if applied in a road safety setting, we will now turn to the role of rewards in models of road user behaviour and review empirical findings of such programs in the area of road safety.

1.4 Rewards and road user behaviour

It can be said that, in terms of the utility point of view, the subjective costs of safe road user behaviour often appear to dominate the subjective benefits. The lack of interest for measures to increase safety is strengthened by people underestimating their chance of being involved in an accident. In addition, people are in general bad at estimating small probabilities; the probability of being actually involved in an accident is very small (Slovic et al., 1982). Moreover, people look for arguments that are consistent with their behaviour; those who do not use safety belts, for example, can adopt the idea that using a safety belt actually increases the chance of serious injury. To make the subjective benefits of safe behaviour greater than the costs, reward campaigns can be employed. At the same time, they are intended

to bring about behavioural changes without people experiencing them as an attack on their personal freedom. Punishments are often seen as such (see, e.g., Elman & Killebrew, 1978; Geller, 1984b; Kunreuther, 1985).

1.4.1 The concept of reward in models of road user behaviour

Many attempts have been made to find unifying principles or models underlying driving behaviour (see, e.g., Evans, 1991; Lonero et al., 1994; Michon, 1985; Rothengatter & Carbonell Vaya, 1997; for overviews), but only a few models have given the concept of reward an explicit role.

Wilde (1985, 1988, 1994; see also Wilde & Murdoch, 1982) has advocated the use of incentives based on his 'risk homeostasis theory' which was derived from a basic economic utility theory that assumes a balance between the risks and benefits of choosing safe or unsafe behaviour (see also Lonero et al., 1994). Risk homeostasis theory has been controversial for various reasons (see, e.g., Evans, 1991; McKenna, 1988). It maintains that, in any activity, people accept a certain level of subjectively estimated risk to their health, safety, and other things they value, in exchange for the benefits they hope to receive from that activity. Individual road users experience or anticipate, at any moment of time, a certain amount of danger, and they compare this with their target (accepted or preferred) level of risk. The accepted level of risk operates as the unique controlling variable of a closed-loop regulated process of which accidents are the output. According to risk homeostasis theory the only way to increase safety is by reducing the accepted level of risk.

The target level of risk can be reduced by interventions in four categories of tactics:

1. decrease the perceived benefits of risky behaviour (e.g., paying taxi drivers per time unit instead of per kilometre);
2. increase the perceived costs of risky behaviour (e.g., penalties for traffic violations);
3. decrease the perceived costs of cautious behaviour (e.g., tax exemptions on safety equipment);
4. increase the perceived benefits of cautious behaviour (e.g., rewards for accident-free driving).

Wilde mainly favours the last tactic. Of all countermeasures that affect people's motivation towards safety, Wilde (1994) considers those that reward people for accident-free performance the most promising. Some examples of this class of

countermeasures are administrative incentives and rewards for accident-free and violation-free driving through discounts in driver license renewal fees, vehicle permits and medical-insurance premiums, and reductions in automobile insurance premiums for accident-free driving.

Another relevant model in this context was proposed by Janssen (1990, 1991; see also Janssen & Tenkink, 1988), who used a utility function model describing total trip cost from the driver's point of view as a function of expected accident costs plus the costs of time loss. Speed is the prominent variable in this model. Janssen suggests that while risk compensation can be seen as the negative side of the adaptability of behaviour to a changing environment, the expected effects of incentives form the positive side. It is assumed that incentives, expressed as a fixed monetary award per kilometre driven without accidents, alter the utility of a trip. On the basis of this utility model, expected safety effects are derived for various incentive schemes. For example, from this model it appears that proportional rewarding (to the actual reduction in accident rates achieved) will always yield better results than all-or-nothing schemes (in which only a reward is given for not having an accident at all). Furthermore, an incentive is assumed to be more effective for longer contingency periods (in time or kilometres driven) and more effective when more individuals as 'members of a group' are involved. In terms of the proposed model, no differential effects of direct versus lottery based incentives are to be expected.

Finally, Fuller's (1984, 1991a) model is considered relevant here. He explains aspects of driving in terms of a threat avoidance model. He uses the term threat to indicate that most of the time on the roadway the driver is dealing with potential aversive stimuli or threats; he uses the term avoidance because for much of the time the driver seems to be either avoiding aversive stimuli (for example, steering around obstructions) or avoiding the possibility of aversive stimuli arising (for example, reducing speed or selecting a clear lane). The driver is not so much trying to avoid crashes, but trying to avoid unpleasant experiences, which in some cases might be precursors to crashes (see Evans, 1991). It is suggested that, when confronted with a discriminative stimulus for a potential aversive event, a driver's response depends on the rewards and punishments for particular alternatives. Because of the conditioning of anticipatory avoidance responses to particular discriminative stimuli, road users have learned specific production rules that generally lead to rewarding choices in that they prevent unpleasant experiences. However, rewarding choices might not always be safe choices. As Fuller (1991b) points out, accidents may also arise because drivers become entrapped by previous experience and the attraction of rewarding outcomes which compete with safety. For example, when unsafe behaviours are rewarding (the 'consequences trap'), when the drivers do not know the contingencies ('contingency trap'), and when the driver has learned unsafe behaviour ('conditioning trap'). Behavioural traps arise because the contingency between for example speeding, which is gener-

ally experienced as a rewarding driving behaviour because it means shorter travel time and more excitement, and a hazardous consequence is improbable and uncertain (cf. Rumar, 1988). An external reward might shift the balance between the perceived pleasant and unpleasant experiences associated with certain behaviours towards the safe direction.

One of the differences between these models seems to be that while risk homeostasis theory (and to a certain extent also Janssen's utility model) accentuates that people choose to maintain a preferred level of risk, the threat avoidance model accentuates that drivers try to avoid unpleasant consequences (one of these might be 'risk') (cf. Brehmer, 1994). As Wagenaar (1990) has pointed out, running risks is not the same as taking risks. Another difference that is of particular relevance in the present context, is that while in the first two models (effective) rewards are assumed to be contingent on a certain outcome (no accidents), in the latter model also rewards contingent on certain behaviours are considered.

Some problems associated with models advocating outcome-oriented rewards are that these often fail to specify what the driver's required behaviour is (what is safe driving?) which makes it difficult to specify precisely what changes in behaviour have occurred once a safer accident record has been achieved (see, e.g., Fuller, 1991b). On the other hand, others (e.g., Wilde, 1985) have pointed out it is better to reward collision avoidance instead of some specific 'safe' behaviour. If a specific behaviour is modified to be safer by an external incentive, and the person's desired level of safety has not changed, then the safety benefit could be offset in favour of some other value, such as saving time. Also, McAfee & Winn (1989) remark that "The risk is there that while the rewarded behaviour may improve, other related safe behaviours may deteriorate" (p.14). If a broader criterion, such as avoiding collisions, is rewarded, than it is reasoned that all the safe behaviours that contribute to it are also rewarded and maintained. However, reward programs intended to modify driver behaviour without specifying the required responses, but rather the required consequence (outcome) can create problems for the driver who does not know how to behave to achieve this consequence; it also creates problems for the researcher who will find it even more difficult to determine the mechanisms through which any observed improvements are mediated (Fuller, 1991b). Bower (1991) pointed to a constraint in the use of incentives, in that the low level of reporting of unsafe acts and even of collisions would reduce the precision of the contingency between an individual driver's performance and reward.

Furthermore, with regard to behaviour-oriented rewards, differential effects of rewards might exist for discrete and continuous behaviours. Examples of discrete ('on/off') behaviours are wearing a safety belt, a helmet, or putting on lights. Examples of continuous behaviours are speed (choice) behaviour, following distance to the next car, and steering behaviour (e.g., lateral position). In earlier sections

it was noted that for rewards to be effective, it is important that the reward contingencies are simple and clear; one should know exactly how to behave and how this is monitored in order to get a reward. In this sense, it can be reasoned that it is easier to influence discrete behaviours by means of rewards than to influence continuous behaviours.

A variety of reward programs have been studied in the area of road safety. These programs can be categorized as either outcome-oriented or behaviour-oriented. A number of studies have investigated the effect of (material) rewards on the outcome of 'unspecified driving behaviour': accidents versus an accident-free driving record, where it is assumed that the driver is aware what concrete behaviour is required to prevent accidents. Other studies have investigated the effects of rewards on the occurrence of specific discrete behaviours (such as using a safety belt) or their influence on continuous behaviours (speed).

1.4.2 Effects of outcome-oriented rewards

Laboratory studies

A few laboratory studies have been conducted that investigated the effects of rewards on subjects' decision making in simulated traffic conditions. Janssen (1988, 1989, 1991; see also Levelt, 1992) found that in an operational game various incentive schemes reduced subjects' accepted risk levels – in terms of the total loss over a series of speed choices – relative to a baseline condition without incentive. The outcome (total loss) was defined as the accumulated hourly losses and 'accident' costs at chosen speeds. Incentives were based on group or individual performance, were either a direct bonus or a lottery ticket, and the extent of the period to be performed damage-free was varied. The incentive's effect was largest (40% reduction in risk level) when it was a direct bonus contingent upon group (six persons) performance. When groups were to earn their incentive by means of a lottery the effect dropped to about 10%. For individual incentives there was not a large difference between the 'direct' and the 'lottery' condition (probability of winning 1/100 with a value 100 times more than the direct reward), although the lottery condition seemed to work somewhat better. Both individual conditions were inferior, in terms of reduced risk, to the condition where groups were directly rewarded for accident-free performance. There was an appreciable gain when the contingency was extended from one to six consecutive damage-free choices. However, no further gain was achieved when the contingency was extended to 12 choices. Janssen concludes that, in general, the findings were consistent with his proposed expected-utility model of decision making (see p. 18–19).

These results do not seem to be in accordance with those of Tränkle & Gelau (1992), although the experiments they conducted are difficult to compare. Tränkle and Gelau used a computer simulation of an intersection situation, and subjects were asked to move as many vehicles as possible across the intersection making use of the gaps in the flow of traffic while avoiding collisions with the crossing traffic. The level of difficulty of the task was varied by changing the number of suitable gaps in the moving traffic. The relative size of rewards for successful crossings and of the penalties for collisions were varied. The results of this experiment showed that subjects reacted to increased or decreased difficulty of task by changing their behaviour in such a way as to keep the frequency of collisions at an almost constant level, seemingly without considering possible benefits (level of reward) and losses; subjects did not earn the maximum possible points.

Industrial programs

The majority of reward programs to promote safe behaviour in traffic have been applied in industrial settings. For example, a survey among Dutch transportation companies indicated that 80 out of 475 (17%) companies contacted run programs involving some sort of reward that are aimed at safe and cost-effective road user behaviour by their employees (Simmelink & Heere, 1993). Businesses seem to provide the most ideal surroundings for the application of such programs. If workers demonstrate safer behaviour on the road, for example, this should reduce absenteeism and (indirectly) increase productivity, as fewer serious accidents occur (e.g., see, Geller, 1984b). In addition, usually a specific, homogeneous population is involved within a generally small-scale organisation, which simplifies practical realization of the program.

The relationship between the costs and benefits of such reward programs generally seem very favourable. Companies which introduced such programs have reported a doubling in cost savings on accidents in relation to the cost of running the campaigns. Those incentive measures that cover a specific period, such as six or twelve months, and use progressively larger rewards for longer periods of accident-free driving, seem to be particularly effective in terms of a reduction in the number of accidents, both per person and per kilometre. Campaigns such as these have been reported to lead to substantial accident reductions of 50 to 80% per worker (Gray, 1990; KRAFT, 1987; Wilde, 1985).

A German company, for example, has been using an incentive program for their drivers since 1957. After each six-month period of accident-free driving, the driver receives a monetary reward. After 36, 60, 120 and 180 months the driver is awarded inscripted golden ornaments of increasing monetary value. From information

provided by the company (KRAFT, 1987) it appears that the number of culpable accidents has decreased from 103 in 1956 to 45 in 1986. The accident rate (culpable accidents per 100,000 kilometre driven) has continuously decreased over the years (from 1.7 in 1956 to 0.35 in 1986) as have the direct accident costs (repairs). Also, Gray (1990) reports that a safety program which involved progressively larger rewards for one, three and five years of accident-free driving, respectively, reduced the accident rate of company cars from 64% in 1983 (87 accidents with an average fleet of 137 cars) to 35% in 1989 (67 accidents with a fleet of 184 cars).

Reward programs based on group performance and team competition have also shown to be effective in reducing accidents, although negative side effects have been reported as well. For example, programs in which rewards were offered to the team of employees with the fewest accidents have shown to be effective in reducing accidents (see, e.g., Baum & Kling, 1997; Caulkins, 1971; Haynes et al., 1982, cited in McAfee & Winn, 1989). However, a questionnaire study among employees of a company that ran a group based reward program has shown that more than half the employees regarded the games and prizes related to the campaign as childish. In addition, more than three-quarters of the employees indicated that not all accidents were reported (Boonstra, 1985; Boonstra et al., 1982). Mixed results have also been reported by Sagberg (1994), who evaluated a safety program (a.o.) involving group based monetary rewards of increasing values for larger accident reductions. In one (large) department of a Norwegian company a decreased accident rate was found as compared to a control group. However, in another (small) department an actual increase in accident rate was found. The authors suggest that an explanation of these results might be the considerable organizational change in this department at the time of study. The main conclusion of the authors is "that there seem to be some – although rather weak – indications that the implemented activities may be effective".

In conclusion, industrial programs involving rewards have shown to have substantial effects in terms of accident reduction. However, it is difficult to determine the exact size of the effects of such programs. First, it is often not clear to what extent the programs have led to the under-reporting of accidents (in order not to lose the reward). Second, evaluations of these programs did not always have adequate research designs (e.g. control groups are often not included), and results are often based on company reports. Finally, rewards are usually embedded in safety programs involving (many) other components as well, making it impossible to disentangle the effects of different components. In one study, for example, the effects of various safety programs to reduce injuries and costs within a Swedish company were compared (Gregersen et al., 1996; see also Gregersen & Morén, 1990). Programs included driver training, group discussion, publicity efforts, and rewards for accident-free driving (no-claim bonus). The results of this study show that although all programs succeeded in decreasing the accident rates compared

to a control group, group discussions and driver training were more effective than rewards, and group discussion was the most cost-effective.

Large-scale accident studies

The first larger-scale application of incentives was carried out in California as part of a driver improvement program (Harano & Hubert, 1974; see also Lonero et al., 1994). Drivers who had caused collisions or committed violations in the previous year were informed by letter that they would receive a free 12-months extension to the validity of their driver's license if they maintained a clean record during the forthcoming year. Apart from the (small) financial incentive, this offer also implied deferral of the written driver's examination, which was required for license renewal. In the first follow-up year, significantly fewer drivers in the incentive group (as compared to a control group) had collisions. The effect was strongest for those drivers whose license renewal was to come up within one year after receipt of the letter. In this group the collision rate was 22% lower than in the comparable controls. The drivers who actually earned the benefit after one year had 33% fewer collisions in the second follow-up year than did the controls. Another group of drivers was given the free license extension unexpectedly after a one-year period of clean record rather than being offered in advance as an incentive. These drivers actually performed worse than controls in the subsequent period. An explanation for these differential effects may be that an unexpected reward without the prospect of a future reward for accident-free driving does not motivate drivers to purposely change their driving style; they did not 'earn' the reward, but 'accidentally' received it.

Other variations of offering license renewal by mail as a 'reward' instead of in-person renewal to selected groups of drivers with clean records showed no significant differences in effect as compared to controls (Kelsey et al., 1985; Janke, 1990 although these authors do not use the term reward in their study).

Related findings were reported by Marsh (1978), who studied the effects of educational programs for drivers with demerit points on their accident and conviction records. Some of the programs included an incentive (the reward was a one-point reduction from the individual's demerit-point record) for future good driving – i.e. if they went six months without a new entry on their record. Only the group that was promised the reward if they sent in homework and did not commit more offences during the next six months, showed significantly fewer accidents than the control, and that reduction was restricted to the second six months. The authors remark that the benefits were more related to the promise of the incentive than to its subsequent delivery.

Peck et al. (1980) report that an analysis of the so-called traffic violator school (TVS) program showed that the TVS treatment did not have a significant effect during the subsequent six months on accidents or convictions. When 'traffic violators' take part in this TVS treatment their conviction is dismissed. The rationale is that the dismissal of the conviction promotes safe driving by supplying positive reinforcement and an incentive to attend traffic school. Peck et al. argue that this rationale is, at best, dubious. When a driver receives a dismissal for attending traffic school, he or she is, in a sense, being rewarded for attending traffic school. However, what is wanted from a behavioural perspective is to reward safer driving, not merely attending traffic school. The authors conclude that any reward system that is not contingent upon maintaining an improved record may be counterproductive, as was the case for TVS.

A final example is provided by Schaaf & Granderath (1989). They report on a three-year campaign in Germany intended to make young drivers aware of the risks of traffic and the limits of their own ability and skill. Annual prize competitions were one element in the campaign. Schaaf and Granderath report that the results were positive as far as improving the awareness of the public is concerned. However, no empirical data appear to be available on the influence of the reward activities on accident rates (Kreisverkehrswacht Bautzen, 1998).

In conclusion, larger scale applications of rewards have concentrated on drivers with 'poor' driving records who were offered an incentive for future driving with 'clean' records. These studies generally show positive effects on driving records, but negative effects have been reported as well. In particular, 'unexpected rewards' and rewards not contingent on safer outcomes (but on, e.g., attending a course) appear to have adverse effects.

Discounts on insurance premiums

Examples of rewards can also be found in the premium discounts car insurers offer their clients, for example, on the basis of the region where the insured party lives, or sometimes on the basis of gender. These types of discount are not individual rewards in the sense that the discount is determined on a statistical basis for certain population groups (Hagenzieker, 1988). A young male driver, for example, may drive very safely, but his premium remains high for several years because, statistically, he has a greater probability of being involved in an accident as compared to other groups of drivers. Discounts are also applied depending on the number of years of accident-free driving. However, the discounts in insurance fees are usually not progressive with the year-to-year accumulation of an accident-free record. Moreover, additional insurance discounts are usually not extended beyond a limited number of years of accident-free driving. If a driver is at fault in an accident

after this period has lapsed, a 'forgiveness clause' is in effect, which may save the driver from an increase in his insurance fees if he has an accident (Murdoch & Wilde, 1980; Wilde & Murdoch, 1982).

In addition, the danger is that accidents are not reported in order not to lose the premium discount. Furthermore, while such premium discounts may represent an indirect reward for good behaviour, these (postponed) rewards come so late that they can hardly be considered as reinforcers of specific safe behaviour in traffic (Hurst, 1980). It is also possible that the loss of a premium discount as a result of submitting a claim is more likely to be regarded as a punishment than its retention is regarded as a reward by the insurance holder. Seen from this perspective, a more effective form of reward, for example, would be to repay part of the premium for accident-free driving in a certain period (Bower, 1991; Hagenzieker, 1988).

Finally, most automobile insurance providers are commercial companies that cannot be expected to implement incentive programs with vigour and persistence, because a very low accident rate is not in their business interest. This (a.o.) led Wilde (1988) to conclude that the objective of road safety promotion would be best served if the responsibility for automobile accident insurance is assigned to the public sector. An example of a government-run automobile insurance institute can be found in Canada (Gélinas, 1992; Outreville, 1984; cited in Baum & Kling, 1997).

Empirical results on the effect of insurance incentives on accidents are virtually absent. At least, these are generally not reported in the (scientific) literature (see Baum & Kling, 1997; Mulder & Wesemann, 1993; Perkins, 1990; Twisk & Mulder, 1992). Since most insurers are commercial enterprises, it is hardly surprising that even if results were available these companies would not make such results public. Nevertheless, many assume that variable premiums exert a positive effect on road safety. However, Bijleveld (1998) recently concluded that bonus-malus type insurances are not generally associated with fewer accident-related claims than other types of insurance. He compared accident-related claims of groups of Dutch drivers with and without a bonus-malus type insurance. Although analyses of data-files of insurers showed that some groups of drivers had fewer accident-related claims as compared to other types of insurance (which premiums were independent of claim behaviour by clients), for other groups of drivers the reverse was found.

Only one other study was found directly relating insurance incentives to accident claims. Vaaje (1991) describes a study in Norway in which the effects of repaying a part of the premium as a reward after an accident-free period were investigated. In a preliminary evaluation of the incentive effects of the rewarding system conducted from September 1989 to July 1990 (compared with control groups), accident rates (claims per 100 cars) before and after the introduction of the reward

system were investigated; the analysis was based on a total of over 90,000 accidents. The reward group (18–22 year old drivers) had a 35% reduction in reported claims, whereas various control groups all had 12–13% less claims compared with the before-period. Despite of these promising results in reduction of accident rates among different groups of young car owners, Vaaje (1991) states that it is unlikely that the reward system has changed the drivers' behaviour so dramatically that this change is the only or main explaining factor for the improvements in the number of reported claims. Alternative explanations may be the under-reporting of accidents (in order not to lose the reward) and risk selection (certain drivers might have been attracted by the reward system because they consider themselves as 'good risks', whereas 'bad risks' might have left the company after the extra premium is added because they do not think they would receive the reward anyway).

1.4.3 Effects of behaviour-oriented rewards

Continuous behaviours

There is hardly any research available on the effect of rewards on continuous road user behaviours. A few studies relating rewards to speed behaviour of drivers have been found. However, not all of these studies address the influence of rewards on speed directly and the results are not conclusive.

For example, Golob & Hensher (1994; see also Hensher et al., 1991) conclude from a survey among truck drivers that financial rewards, in terms of earning rates per kilometre driven or per trip, have a significant impact on their reported behaviour. With regard to speeding behaviour it appeared that higher earning rates had significant negative influences on average (reported) speed and speeding fines. However, because the study was not set up to experimentally investigate the effects of financial rewards on speeding behaviour (but used 'earning rates' as a background variable to explore correlations with a number of behaviours) and because these findings were solely based on self-reported behaviour of the respondents, it remains to be seen whether positive effects of financial rewards per sé would also become apparent in the actual behaviour of (truck) drivers.

In the frequently cited studies by Van Houten c.s. (e.g., Van Houten & Nau, 1983; Van Houten & Malenfant, 1993), the effects of a combination of feedback and enforcement on speed behaviour were investigated. In some studies also an incentive-component was part of the program. For instance, motorists driving within the speed limit were stopped and thanked for driving within the speed limit (Van Houten et al., 1985). They were also given a small certificate and a pen. The results of this study show that the percentage of drivers exceeding the speed limit

decreased during the interventions. However, the isolated effect of the incentives is not known, because the campaigns always involved other treatments (enforcement) as well. Moreover, feedback signs alone, indicating the overall level of compliance, have shown to have a marked effect on speed choice (see e.g., Van Houten et al., 1985). Rooijers & De Bruin (1991) note that one explanation for the effect of (collective) feedback could be that it motivates drivers to conform to others and probably leads to an implicit notion of surveillance. In this sense, the feedback message would not even have been interpreted by drivers as a 'reward'.

A number of field studies have addressed the effects of rewards on speed behaviour more directly. For example, in a German study (Machemer et al., 1995; see also Wolf & Machemer, 1994; Tücke & Wolf, 1996), the effects of 'delegated rewards' were investigated: drivers were not individually rewarded but rewards were donated to an institution (e.g., kindergarten) in the limelight of public interest when a certain pre-defined target in speed reduction was reached. Machemer et al. (1995) concluded that in one municipality the delegated reward was clearly successful in reducing speeds, but in a second case it was considered as failed. In a third municipality only insignificant speed reductions were achieved and at only a few points. On the other hand, enforcement activities led to clear reductions in speeding rates on a short term basis at all three municipalities investigated.

Another field experiment investigated the effect of a promise to drivers of a considerable incentive (Dfl. 800,=) for not having a culpable motor vehicle accident over a period of a year (Janssen, 1994; see also Levelt, 1992). Although this is an example of an outcome-oriented reward program, it is mentioned here because also several behavioural variables served as dependent variables, e.g., driving speed, acceleration, and time headways to leading vehicles in the same lane. The expectation was that this group would become more 'careful' in their driving, for example resulting in lower mean speeds. However, the effects of the incentive on driving behaviour were found to be largely absent. Apart from a decrease in irregularity in speed in the first follow-up measurements – followed by an actual increase in the last measurement – drivers in this condition did not show significant changes in the behaviour.

The most comprehensive study that has addressed the effects of rewards on speed behaviour is by Heino (1996; see also Heino et al., 1996). He did not specifically ask drivers to drive within speed limits or lower their speeds to receive a reward, but promised them a reward when they would drive 'more safely' as compared to a first test drive for which they were instructed to drive as they 'normally do'. It appeared that drivers decreased their speeds significantly in the incentive conditions as compared to control conditions. Two incentive conditions were distinguished: a direct reward of Dfl. 10,= was promised if they would succeed in driving more safely, or a lottery ticket with a chance of winning a prize (Dfl. 40,= to Dfl. 100,=).

The results show that both conditions resulted in a lower mean speed in the second test drive as compared to the first, while no differences in mean speed were found for drivers in the control condition. The largest speed reduction was found in the lottery condition. It was also concluded that lottery incentives had a larger influence on those drivers who tend to be fast-driving people ('sensation seekers').

Discrete behaviours

Only a limited number of discrete behaviours have been the topic of empirical investigations studying the effects of rewarding these behaviours.

Some studies have addressed whether rewards for *sober driving* could be effective. However, rewarding people who drive a car while not under the influence (of alcohol) can be associated with some major problems. For example, the 'breathalyser' equipment that the police use is often not able to determine someone's exact blood alcohol level. Someone who can light up the pass-light on the breathalyser and receive a reward on this basis might either have drunk nothing at all or – under certain circumstances – five to six glasses of alcohol. To reward a person in the latter case can stimulate undesired behaviour. (It should be noted that this argument also applies to enforcement: one can commit an offence of consuming too much alcohol, and remain unpunished despite being checked by the police.) In addition, people who, after consumption of alcohol, have deliberately chosen not to drive are excluded if only sober drivers receive a reward (cf. Mathijssen et al., 1989).

A few examples of studies on rewarding drivers for sober driving can be found in the literature. However, the effects on actual 'drinking behaviour' of such programs are virtually unknown, either because evaluations were based on indirect measures such as questionnaire results showing public support of such a program (Caverson et al., 1990), because the programs consisted of combinations of interventions which do not allow effects to be contributed to the isolated elements (Mathijssen, 1992), or because the programs did in fact not reward sober driving, but instead rewarded taking part in a treatment program (Sadler et al., 1991) or rewarded the organizing of such programs (Overend, 1984). Only one study involved behavioural effects of an 'isolated' incentive program (not including other interventions such as enforcement) more or less directly (Brigham et al., 1995). However, also this study included no direct measurements of actual drinking behaviour.

Mathijssen (1992) evaluated a three-month enforcement campaign that also involved a reward component. Drivers who had put a particular sticker on their car window were eligible for a prize if they passed a blood alcohol test when stopped

by the police. Contrary to testing procedures in previous campaigns (see Mathijssen et al., 1989), this time testers were used that displayed the exact blood alcohol content of the tested drivers, which prevented that drivers could be rewarded for 'sober' driving when they had in fact consumed (a small amount of) alcohol. After the end of the campaign, alcohol use by car drivers appeared to be at a slightly lower level as compared with the before situation. It is, however, impossible to deduce the possible effect of the reward component from the (small) effects of the campaign as a whole.

Brigham et al. (1995) evaluated the effects of a small-scale campaign consisting of prompts and incentives on the frequency of designated driving at a bar frequented by undergraduates. In a before and after study the bar fluctuated between its regular and an enhanced program for designated drivers, i.e. those who indicate not to consume alcohol because they are going to drive after visiting the bar. Observations (the observers were able to locate a total of 194 of the 209 designated drivers identified to the bartender) showed that the enhanced program resulted in an approximate doubling of the number of designated drivers (from a median of three to a median of seven per night).

Although some have attempted – but not succeeded – to encourage the *use of bicycle lights* by reward (Ferguson & Blampied, 1991) or have expressed ideas to reward the *use of helmets* by moped riders (Mathijssen & Verhoef, 1992), the great majority of studies involving discrete behaviours have been directed at increasing safety belt use by car inhabitants.

Effects of rewards on safety belt use

In the past 15–20 years a large number of reward programs to stimulate safety belt usage have been implemented and evaluated, most of them in the USA. Their initial application appears to have been inspired by the absence of mandatory safety belt use laws in many states at the time – excluding enforcement activities as a possibility to influence use rates. More recent reward programs have been conducted to investigate their effect as an alternative for or in addition to enforcement strategies to increase safety belt use. For example, Kalsher et al. (1989) conducted a study at two large military bases in Virginia, where it is compulsory to wear a safety belt. At one base, an incentive campaign was organized, while at the other, the campaign consisted of increased police enforcement. In this way, the effect of a reward strategy could be compared to the effect of a different, more traditional measure. The researchers found that both campaigns realized an equivalent increase in safety belt use. Mortimer et al. (1990) found that the greatest effect was demonstrated when reward and police enforcement were applied in combination.

It should be noted that the initial level of compliance to safety belt use in this latter study was quite low (25–30%).

A number of relevant factors that have been identified in previous sections of this chapter can also be found in the studies on the effects of rewards on safety belt use; for example, the setting of the program, the size and frequency of rewards, and whether rewards are individual or group based, or delivered as a direct reward or in the form of a lottery ticket. Most of the reward programs were conducted amongst relatively small, homogeneous groups, for example company employees, university personnel, or were confined to a particular city or region. Generally, rewards are given on the basis of actual observed use of the safety belt. The rewards varied from certain privileges in the working environment, such as extra time off, small gifts and gift coupons to the chance of winning a prize or gaining social attention, such as ones name or picture in the paper. Most rewards for safety belt use were presented immediately upon observation; mostly in situations where cars have to stop anyway, for example at the point where workers enter an industrial complex. Sometimes rewards in the form of a lottery were offered, providing a chance to win a prize. In this case, for example car number plates of safety belt users were noted down and selected, and the owners of the winning number plates could then collect their prize (Geller, 1984b). Sometimes both immediate and delayed rewards have been used (e.g., Campbell et al., 1984). Also rewards have been used that were not directly contingent upon actual safety belt use. For example, all car drivers who had signed a promise to wear their safety belt were eligible for rewards (e.g., Horne & Terry, 1983b). The value of rewards that are offered in the various campaigns vary a lot: from pens, luncheon vouchers, and small amounts of cash, to cars; and the probability of receiving a reward (co)varies from almost 1 for small rewards to close to zero for very large rewards. All such programs roughly led to a doubling in the percentage of safety belt users. It should be noted in this context that until some years ago, it was not compulsory to use a safety belt in most American states, and hence the usage percentages were very low: often not more than about 10%.

Like methods of individual reward, it appears that rewards which depend on group performance are often successful (see, e.g., Cope et al., 1986b; Geller & Hahn, 1984). In such programs, for example, the level of reward was dependent on the level of the average safety belt use in a group; the higher the group average, the greater the reward. In general, safety belt usage declined considerably once the incentive campaign had stopped, but in most cases, the use of safety belts still remained higher than prior to introduction of the campaign. Research performed by Geller (1983b) demonstrated that the more often motorists receive a reward, the more often they use their safety belts during the follow-up period.

A problematic aspect in reviewing the results of individual studies on rewarding safety belt use is the poor systematic variation in campaign characteristics and settings. A first inspection of the results seems to indicate that most campaigns work, irrespective of their exact form or content. However, not all incentive campaigns to stimulate safety belt use are successful. For example, Foss (1989) reports that a community-wide incentive program aimed at increasing safety restraint use among young children was minimally effective. As the various forms of rewarding were never systematically varied, it is not possible to draw any conclusions regarding the separate effects of program characteristics.

Insurance incentives for safety belt use

In addition to (outcome-oriented) incentives in the form of premium reductions or refunds for accident-free driving, insurance companies could also offer incentives for specific (safe) behaviours, such as using a safety belt. However, there are hardly any empirical results available on the effects of such insurance incentives, and insurance companies, in general, do not appear to be willing to introduce such incentives, because there are serious problems with trusting the insured to wear belts and with denying payments to insured who are injured while not wearing belts (Coonley & Gurvitz, 1983). In addition, it has been suggested (see, e.g., Orr, 1982; Wilde, 1994) that insurance premium discounts should not be given for specific behaviours such as safety belt use or the purchase of safety devices alone, but that discounts should also be dependent on the final outcome (not having accidents).

Robertson (1984) reported on an American insurance company that in 1983 raised the payments to insurance holders who were injured in a motor vehicle whilst using a safety belt. A survey into reported safety belt use, performed one month after the insured parties were informed about these increased payments, indicated that drivers who were insured by this insurance company reported the same degree of safety belt use as those drivers insured by other companies (9 vs. 13%, respectively). While data were not available on belt use prior to the increased incentive, given the low reported belt use among the drivers insured by this company, Robertson remarks that it is unlikely that belt use was even less prior to the announced increment. It is therefore concluded that the incentive of increased payments appeared to have had no apparent effect on (reported) safety belt use.

1.4.4 Conclusion

In only a few models of road user behaviour the concept of reward is given an explicit role. These models imply that external rewards could be (major) motivators to make drivers to change their behaviour in the desired – safe – direction. It is also assumed that some (unsafe) behaviours, such as speeding, are intrinsically rewarding; to make the subjective benefits of safe behaviours greater than the subjective costs external rewards could be applied. While some models stress that rewards should be outcome-oriented, it can be argued that it is also necessary to specify what behaviour is required from road users to attain that outcome.

Research has shown that both outcome- and behaviour-oriented rewards applied in various forms and in various settings can be effective in reducing accidents and changing road user behaviour in the desired direction. However, also negative side effects have been reported such as an under-reporting of accidents. When rewards are given unexpectedly (rather than that these could be anticipated by the drivers) even adverse effects have been found. Rewards dependent on group behaviour have shown mixed results. Hardly any research is available on the effects of rewards on continuous behaviours. Some studies investigating the effects on speed behaviour show no conclusive results. Programs aimed at rewarding discrete behaviour appear to be particularly effective. Much of the work studying the effects of rewards on discrete road user behaviours is inspired not so much by models of road user behaviour, which are not very specific in how to design reward programs, but by principles derived from learning theory, behaviour modification and behavioural engineering. However, research in this area has mainly concentrated on stimulating safety belts by rewards, and it is not known whether rewards for other discrete behaviours will be effective as well. In addition, due to the poor systematic variation in reward program characteristics and settings, it is not possible to determine which (combinations of) factors are optimally effective.

1.5 Rewards and road user behaviour in retrospect

Modifying road user behaviour by reward has not been given much attention in the practice of road safety as well as in models of road user behaviour, despite the fact that a variety of more general theories of behaviour change reserve an important place for influencing behaviour through rewards. In addition, rewards have been successfully used in other fields of application. From the various approaches it appears, for example, that reward characteristics, type of setting, and type of behaviour can be identified as important factors mediating the effects of rewards on behaviour. A review of empirical findings in the area of road safety shows that

both outcome-oriented and behaviour-oriented reward programs can be effective, that such programs generally appear to be viewed positively by the participants, and that these can also be cost-effective from a financial point of view.

Some problems are associated with the application of outcome-oriented rewards. For example, these can lead to an under-reporting of accidents. Outcome-oriented reward programs can also create problems for road users who do not know how to behave to achieve this outcome, and when an accident reduction is achieved it is difficult to find out which specific behaviour change has led to a reduction in accidents. In addition, most (psychological) theories indicate that rewards should be contingent on specific behaviours, and not merely on outcomes. On the other hand, the danger of rewarding some specific behaviour may be that while the rewarded behaviour improves, other related safety behaviours may deteriorate. However, such behavioural adaptation mechanisms cannot generally be assumed to apply to all road user behaviours and safety measures (see OECD, 1990; Grayson, 1996). There are several unresolved issues regarding the crucial factors determining whether, and to what extent, a safety measure results in behavioural adaptation (see, e.g., Grayson, 1996; Sagberg et al., 1997). In the case of safety belt use, for example, it has even been reported that unbelted drivers are more often involved in accidents than belted drivers (e.g., Evans, 1987; see also Evans, 1996), and that non-use of safety belts predicts other risky behaviour to a certain extent (and not the reverse; see, e.g., Mäkinen et al., 1991). It was therefore decided to – in the first instance – further investigate the effects of behaviour-oriented rewards on road user behaviour.

Of all road safety applications of behaviour-oriented reward programs, those directed at discrete behaviours appear to be the most susceptible for finding substantial effects. Therefore, it was decided to first concentrate further research on the effects of rewards on discrete behaviour, and more specifically on the effects of rewards on safety belt use and the factors that mediate the effectiveness. As became apparent in the first part of this chapter, reward programs can be assumed to be particularly effective for this type of behaviour: Rewards seem to be especially effective in modifying (routine) behaviour that is perceived as relatively 'uninteresting' to do, qualifications which appear to apply to the use of safety belts. It has also been suggested that rewards may help 'breaking habits'; the use of safety belts appears to rely much on habitual behaviour (see, e.g., Mäkinen et al., 1991). In addition, rewards have been suggested to be more effective with regard to behaviour that is already known and 'easy to do' (such as putting on a safety belt) as compared to 'new' and complex behaviours (such as adapting one's speed to the situational demands). Also from a practical point of view it is interesting to focus on safety belt use, because this behaviour is relatively easy and inexpensive to measure and monitor.

1.5.1 Research questions and overview of remaining chapters

It can be concluded from the available literature that rewards generally 'work' to (substantially) increase safety belt use. A first inspection of the results suggests that most campaigns conducted in the past have been effective, seemingly irrespective of their exact content; these roughly led to a doubling in the percentage of safety belt use. However, many questions remain. For example, most reward programs have been applied with very low baseline levels of safety belt use, in situations where it was not yet compulsory to use a safety belt. The question arises whether such programs could also be effective in increasing compliance in situations where safety belt use is mandatory (and with relatively high baseline levels), as is to date the case in most (motorized) countries. In these circumstances, for example, police enforcement might be a more effective measure. In addition, not all reward activities have been shown to be successful. Furthermore, it is not clear exactly which aspects of such reward programs mediate the effectiveness. For example, the influence of variables such as the size and type of the reward, the probability and timing of rewarding, and the setting of the reward program (e.g., the role of groups, small or larger-scale settings) on the effectiveness of these programs is not clear. As the first part of this chapter showed, a number of theories consider these variables to be important mediators for effectiveness. It also became apparent in this chapter that external rewards can have adverse effects on intrinsic motivation. However, it is not known whether a reward is perceived as an important (or main) motive for safety belt use by those who have been confronted with reward programs for safety belt use.

In the following chapters it will therefore be attempted to provide some answers with regard to the following questions:

- Can reward programs be effective in raising already relatively high baseline levels of safety belt use?
- Which forms of reward programs are more effective than others – and in comparison to police enforcement – to increase safety belt use with relatively high baseline levels?
- How do the results of such reward programs relate to previously reported results obtained under conditions of relatively low baseline levels of safety belt use?
- Which variables with regard to reward characteristics, settings, and other circumstances can be determined that mediate the effectiveness of reward programs that have been applied to stimulate safety belt use?

A number of studies were conducted to answer these questions. First, a small-scale reward program was evaluated, which was set up at some locations in the province of Friesland in the Netherlands to explore whether rewards would also be effective in raising already relatively high baseline levels of safety belt use. This study is described in Chapter 2. A more elaborate field study was conducted at twelve military bases in the Netherlands, to investigate which of various forms of reward programs would be more effective than others as well as in comparison to various levels of 'punishment' (enforcement). This study is described in Chapter 3 (behavioural observations) and Chapter 4 (questionnaire). The questionnaire of Chapter 4 a.o. contained items concerning motives for using safety belts; for example, to what extent a reward is considered a motive for respondents to use their belt. Furthermore, a meta-analysis was performed to analyse these and previously reported results and other variables of reward campaigns to promote safety belt use. As stated earlier, a problematic aspect in reviewing the results of individual studies is the poor systematic variation in campaign characteristics and settings. By choosing a meta-analytic approach instead of a more traditional, narrative literature review it was expected that this problem could, partly, be overcome. The results of this analysis could also provide some insight in whether the results obtained in Chapters 2–4 can be seen as 'representative' for the effects of reward campaigns obtained elsewhere. The results of this meta-analysis are described in Chapter 5. Finally, in Chapter 6, the main findings are summarized and conclusions are drawn.

Chapter 2

Evaluation of a small-scale incentive campaign to promote safety belt use: A pilot study*

A small-scale incentive campaign to promote safety belt use was evaluated. The campaign was organized in four towns in the province of Friesland in December 1988. Behavioral observations were made to measure safety belt use by drivers and front seat passengers before and directly after the campaign. Available literature on the subject indicated that such an incentive campaign would produce positive results. However, in this case, hardly any effect was found. Possible explanations for these results are discussed.

2.1 Introduction

Rewarding road users to promote safe behavior in traffic is rarely applied. It could serve as an alternative for other measures, such as enforcement. Studies conducted in the USA on the effects of incentive programs to promote the (voluntary) use of safety belts have shown good results (see for an overview, e.g., Hagenzieker, 1988). Based on these findings, it could be expected that even a small-scale campaign with small rewards would result in observable increases in safety belt use. Therefore, a pilot-study was conducted to investigate the effect of such an incentive program on safety belt use in the Netherlands. A small-scale campaign to promote

*This chapter was based on a previously published SWOV report (R-89-20; Hagenzieker, 1989)

safety belt use was organized in December 1988 in four towns in the province of Friesland.

Contrary to previous campaigns, no enforcement by the police was involved, which allowed us to investigate the effects of a 'pure' incentive campaign. Because the safety belt use in the province of Friesland was relatively high, as compared to other provinces in the Netherlands and to use rates in the USA, it was difficult to estimate the magnitude of the expected effect beforehand. Studies conducted in the USA have usually reported a doubling of the safety belt use rate as a result of incentive programs (see, e.g., Geller, 1984b). The fact should be taken into account, however, that in most states safety belt use was not mandatory at that time.

2.2 Method

2.2.1 Participants and setting

At three locations in each of the four participating towns (Leeuwarden, Sneek, Wolvega and Heerenveen) observations of safety belt use were made. In each town one location was located at the entry of a parkinglot of a shopping centre. In addition, two other locations were selected per town, each located within the built-up area. In all cases but one (Wolvega, location 3), locations were situated at cross-sections with traffic lights. Participants were drivers and their front seat passengers in passenger cars at these locations. Delivery vans or cars with covered rear windows (e.g., commercial vans) were excluded from the observations.[†]

2.2.2 Procedure

Observation procedure Baseline use rates were observed in November 1988. Directly after the campaign ended another series of observations were made (December 1988). Both series of observations took place on Saturdays between 10 a.m. and 4 p.m.; two hours at each location in each town.

The observations were executed by trained observers. They were instructed to observe safety belt use by drivers and front seat passengers when they stopped for the red traffic light; when traffic density was low also very-slow-driving cars were observed approaching the traffic lights. The observers scored whether the front seat occupants were using a safety belt, and whether they were male or female.

[†]These types of cars are excluded from the obligation to be equipped with safety belts.

General procedures The campaign was announced in several local and regional newspapers about a week prior to the campaign, in the first week of December 1988. The purpose and contents of the incentive campaign was described in these announcements, namely that the use of safety belts would be rewarded with a small gift in order to stimulate safety belt use. It was not made public where or when exactly the gifts would be delivered, nor what the gift would be. Small rewards (pastry-bars) and an accompanying leaflet on the subject of safety belts were handed over to drivers, who were observed wearing their safety belt when they entered parking lots of shopping centres in these towns. The leaflets were also handed out to those front seat occupants who did not use a safety belt. The rewards were delivered only once by members of the traffic safety organization in Friesland, who were clearly recognizable ('VVN promotion teams'). The rewards were delivered on a Friday or Saturday afternoon between 1 and 5 p.m. A total of 2,500 pastry-bars were handed out; 500 in Leeuwarden, 950 in Heerenveen, 450 in Wolvega, and 600 in Sneek.

2.3 Results

A total of 11,032 observations were taken of which 5,483 occurred during baseline and 5,549 immediately following the campaign. Table 2.1 shows the numbers of observations and percentages front seat occupants using a safety belt for the two series of observations at each of the locations in the four towns. The number of observations was sufficiently large to detect relatively small increases in safety belt use. However, as Table 2.1 illustrates, in none of the four towns and at none of the locations (significant) increases in safety belt use were determined when comparing before and after campaign measurements. Overall mean safety belt use was 63.8% during baseline observations, and 62.9% directly following the campaign. In only one case a significant increase in belt use was found: among female drivers in Wolvega observed at location 3 the belt use increased from 72% to almost 84% ($\chi^2 = 5.18; df = 1, p < 0.05$); all other differences were statistically not significant.

Table 2.1 also shows that the observed safety belt use varies for different towns and locations within towns. In general, at locations situated in the towncentres (e.g., location 1 Sneek) safety belt use is lower than at location near the edges of the built-up areas (e.g., location 3 Sneek). Finally, Table 2.1 shows that female drivers use their safety belt more often than male drivers, a difference of about 10 percentage points; and that front seat passengers – and in particular male passengers – use their safety belt more often than drivers. These observations are in agreement with national findings (Varkevisser & Arnoldus, 1989).

Table 2.1: Observed safety belt use before and directly after the campaign. Results (number of observations N and % safety belt use) are shown for each town and location, and for both male and female drivers and front seat passengers.

	BEFORE CAMPAIGN										AFTER CAMPAIGN									
	MALE				FEMALE						MALE				FEMALE					
Town	Driver		Passenger		Driver		Passenger		Total		Driver		Passenger		Driver		Passenger		Total	
Location	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Leeuwarden																				
Location 1	268	63.1	66	65.2	69	85.5	112	73.2	515	68.5	232	65.0	67	77.6	84	82.1	92	68.5	475	70.5
Location 2	202	63.9	18	77.8	49	73.5	139	69.1	408	67.4	168	66.1	41	56.1	58	70.7	115	83.5	382	67.5
Location 3	170	65.9	18	88.9	75	72.0	139	75.5	402	71.4	169	66.3	16	75.0	72	81.9	127	73.2	384	71.9
Subtotal									1325	69.1									1241	70.0
Sneek																				
Location 1	359	45.4	51	62.7	47	44.7	143	57.3	600	49.7	337	40.1	39	53.8	71	47.9	135	65.9	582	47.9
Location 2	136	50.1	26	57.7	42	57.1	44	75.0	248	56.5	150	52.0	26	57.7	35	54.3	52	59.6	263	54.4
Location 3	18	61.1	2	100.0	2	100.0	10	90.0	32	75.0	121	65.3	27	44.4	34	94.1	56	80.4	238	70.6
Subtotal									880	52.5									1083	54.5
Wolvega																				
Location 1	261	52.5	21	61.9	93	62.4	108	61.1	483	56.7	287	43.9	36	63.8	81	64.2	110	59.1	514	51.8
Location 2	213	65.3	26	73.1	56	62.5	105	87.6	400	71.3	187	67.4	28	71.4	60	63.3	97	75.3	372	69.1
Location 3	449	65.0	78	76.9	118	72.0	252	84.5	897	73.9	437	66.4	72	72.2	136	83.8	249	81.9	894	73.8
Subtotal									1780	68.6									1780	66.5
Heerenveen																				
Location 1	274	51.1	27	55.6	77	77.9	130	73.1	508	61.0	269	54.6	36	61.1	107	70.1	138	71.7	550	62.4
Location 2	263	59.7	46	65.2	109	77.1	135	70.4	553	66.2	246	51.6	33	66.7	97	59.8	123	53.7	499	55.7
Location 3	235	48.1	38	47.4	64	64.1	118	55.9	455	52.3	200	48.5	34	64.7	59	66.1	103	68.0	396	57.6
Subtotal									1516	60.3									1445	58.8
TOTAL	2848	57.2	417	66.4	801	66.2	1435	72.1	5501	63.8	2803	56.5	455	65.1	894	70.5	1397	72.5	5549	62.9

2.4 Discussion

Regardless of town and location, no significant increase in safety belt use was observed when before and after measurements are compared. An exception is the safety belt use by female drivers observed at location 3 in Wolvega, which increased significantly. No explanation can be put forward for this finding; it seems likely that this occurred by chance. The mean observed safety belt use rate found in this study is the same as the mean national belt use rate in 1988 for locations inside built-up areas (Varkevisser & Arnoldus, 1989); 63% in both cases. In 1987 the mean national safety belt use within built-up areas was 55% and 54% in the province of Friesland (HTS Leeuwarden, 1989), whereas in 1985 – immediately following a widescale campaign consisting of a combination of police enforcement and publicity – safety belt use in Friesland was much higher than the national average, namely 65% versus 49% for locations within built-up areas (Gundy, 1986). It seems that, because since 1985 many regional and national campaigns have taken place to increase safety belt use, the 'headstart' of Friesland has in the mean time been caught up by the other provinces.

Available literature indicated that an increase in safety belt use was to be expected as a result of an incentive campaign, even when conducted on a small scale like this (see, e.g., Geller, 1984b; Hagenzieker, 1988). In general, only short-term effects have been reported that had disappeared after several weeks, which made repeated campaigns necessary to obtain longer lasting effects. The observations in the four towns in Friesland indicated that even such a short-term effect did not occur. The question arises why no increase in safety belt use was found in this case.

An important aspect of incentive programs is generally considered to be that prior to the campaign the public should be aware of the planned campaign; it should be known that when behaving in a certain way, drivers can be rewarded for doing so (cf. 'expectancy-valence' theories; see, e.g., Bandura, 1986). There are even examples of a negative effect of rewards when drivers were rewarded unexpectedly (Harano & Hubert, 1974), i.e. in this case they were more often involved in accidents as compared to controls. The collected announcements of the present campaign give the impression that the incentive program was not sufficiently known by the public; the absolute number of announcements was small, and in each local or regional newspaper the advertisement appeared only once.

Another possibility is that rewarding only once is simply not enough. The importance of repeated interventions has been stressed often in the literature (see, e.g., Geller, 1984a). In addition, the reward should be experienced as attractive in order to be able to change behavior (Williams, 1980). From the announcements in the local and regional press can merely be derived that one could get 'a gift'

that ‘is appropriate to the month of December’; the exact nature of the gift was not made public. It might have helped to announce that pastry-bars could be won. On the other hand, even then these gifts could have been considered not attractive (enough) in order for people to change their behaviour.

Finally, it should be mentioned that this pilot-study lacked a control group. Moreover, no follow-up observations of safety belt use were made after withdrawal of the program. However, because the results indicate no increase on safety belt use even directly after the campaign ended, it is felt that this can only in theory be criticized. It is hardly imagineable that on control sites belt use would have dropped and have continued to decrease during follow-up, in which case the incentive program would have had an effect after all.

Chapter 3

Enforcement or incentives? Promoting safety belt use among military personnel in the Netherlands*

During a nationwide campaign to promote safety belt use among military personnel, a field study was conducted at 12 different military bases in the Netherlands. Amount of enforcement, type of publicity, and incentive strategies were varied among military bases. Observations of safety belt use among servicemen in their personal vehicles were conducted: before the campaign, immediately following the campaign, and three months later. Safety belt use increased from 65% during baseline to 73% directly after the campaign, and to 76% three months later. An overall 28.6% increase in safety belt use (from 63% to 81%) was observed at seven bases, whereas no changes were found at five bases (68% on all occasions). To a large degree the effects were due to a 37.7% increase among young drivers. These results confirmed that enforcement, as well as incentives, can be effective in promoting safety belt use. However, treatment effects were not systematic, thereby complicating the interpretation of the results. Implications of these varied outcomes are discussed.

*This chapter was previously published in Journal of Applied Behavior Analysis, 1991, 24(1), 23–30 (Hagenzieker, 1991); reprinted with permission from JABA.

3.1 Introduction

Even though most drivers recognize that vehicle safety belts are effective in reducing or preventing driver injuries, many do not use safety belts. In the Netherlands, a safety belt mandate was enacted in 1975; national belt use rates increased from 20% in 1974 to 50% in 1975 and to 70% in 1988 (Varkevisser & Arnoldus, 1989). However, young male drivers use their safety belts less often than other groups. Thus, legislation alone is not sufficient to achieve universal use of vehicle safety belts. Young male drivers are also frequently involved in accidents; thus safety belt use is especially critical for this group (Van Kampen, 1988). Therefore, a nationwide campaign was designed to promote safety belt use among military personnel, a population made up primarily of young males.

Previous studies have demonstrated that campaigns consisting of a combination of enforcement and publicity can increase safety belt use rates substantially (e.g., Jonah & Grant, 1985; Williams et al., 1987a,b). Gundy (1988) and Grant (1989) each found that after two years, belt use rates remained higher than the original baseline levels.

Incentive programs have been successful in increasing safety belt use (e.g., Elman & Killebrew, 1978; Cope et al., 1986b; Geller, 1988; Geller et al., 1989, 1987). These studies have varied the type and frequency of rewards for belt use, and all have increased safety belt use significantly. However, these findings must be tempered by the fact that all were carried out in the absence of a safety belt use mandate; therefore, baseline use rates in each of these programs were relatively low (10% to 20%). Unfortunately, use rates have typically decreased within a few weeks following withdrawal of the incentive programs. Cope et al. (1986b) presented one of the few studies finding no decrease in postintervention use rates after six months.

Little is known about the relationship between enforcement levels and safety belt use. A direct comparison of the relative efficacy of enforcement and incentive programs has been carried out only once under conditions of mandatory requirement of safety belt use (Kalsher et al., 1989). These researchers evaluated safety belt promotion campaigns on two US naval bases. In their study, the enforcement ('disincentive') program led to greater overall increases in belt use than did the incentive program. Six months after the intervention programs were removed, belt use declined on both the 'incentive' and the 'disincentive' bases. The present study varied systematically, across military bases, the amount of enforcement, type of publicity, and type of incentive strategy.

3.2 Method

3.2.1 Participants and setting

A total of 12 different army, navy, and air force bases located throughout the Netherlands participated. Base populations varied from approximately 400 to 3,500 (see Table 3.1). Participants included all personnel (drafted, enlisted, and civilian) who drove through the entrance/exit gates.

3.2.2 Procedure

Observation procedure Baseline belt use rates were observed June through September 1988; the next measurements took place at the beginning of December 1988, directly after the campaign ended; follow up observations occurred in February 1989.

Observation methods included both unobtrusive observations and obtrusive observations. During unobtrusive observations, a trained observer stood next to the gate and recorded the shoulder[†] belt use of all passing drivers at a distance of approximately 1 m. During this condition, shoulder belt use or nonuse could be determined in at least 99.5% of all observations.

This method, however, did not allow for a distinction between different age groups, types of personnel, or base assignments. Therefore, obtrusive observations were conducted in which military police stopped vehicles before they entered the gates during morning arrivals and after they exited the gates during afternoon departures. Trained observers asked each driver about his base assignment, age, and whether he was drafted, enlisted, or a civilian. At the same time, the observer scored whether the driver was using a shoulder belt. Shoulder belt use could be determined for at least 99.9% of these observations.

Observations took place on weekdays, between 6:30 and 8:30 a.m. for entering vehicles, and between 3:00 and 5:00 p.m. for departing vehicles. For all experimental phases, unobtrusive observations occurred unannounced on a randomly selected weekday, and obtrusive observations took place on two randomly chosen weekdays.

[†]Since June 1, 1975, all passenger cars in the Netherlands (but not, e.g., delivery vans) are equipped with seat belts (lap or shoulder) in front seats; in practice, virtually all passenger cars have shoulder belts. Vehicles without a shoulder belt for the driver were not included in the observations.

Table 3.1: The experimental design. The rows represent the experimental conditions. The first column shows the activities during the first month of the campaign, the second shows the activities during the second month, the third shows the 12 different bases allocated to the conditions, the fourth shows the estimated size of the population at that site, and the fifth and sixth show the number of hours spent on surveillance by the military police and the number of registered fines, respectively.

October 1988	November 1988	Base	N on base	Hours	Fines
no extra publicity	minimal enforcement	(1) P/MI	1800	16	20
extra publicity	minimal enforcement	(2) EP/MI	2200	26	4
no extra publicity	moderate enforcement	(3) P/MO	1500	72	5
		(4) P/MO	850	32	15
extra publicity	moderate enforcement	(5) EP/MO	400	–*	–*
		(6) EP/MO	1400	32	2
no extra publicity	intensive enforcement	(7) P/IN	3500	60	100
extra publicity	intensive enforcement	(8) EP/IN	1850	24	6
extra publicity	group-dependent incentives	(9) EP/GR	600	–	–
extra publicity	group-dependent incentives	(10) EP/GR	600	–	–
extra publicity	individual incentives 1 prize/week	(11) EP/INC1	1000	–	–
extra publicity	individual incentives 4 prizes/week	(12) EP/INC4	800	–	–

* No data available

General procedures The campaign, which began in October 1988, was conducted over a period of two months. The first month was used to announce the campaign. For those bases assigned to incentive treatments, special brochures were disseminated that outlined the intervention condition. Personnel at those bases assigned to enforcement treatments received brochures containing information about penalties for not buckling up.

Other publicity materials included stickers, playing-cards, and posters with the special campaign logo printed on them. In addition, local newspapers, military newspapers, and radio stations provided some media coverage about the campaign.

Extra publicity was used on some bases. For instance, a movie, entitled *An American* was made that portrayed the importance of the safety belt embedded in a thriller-like plot about a boy and his 'old American dream car.' Special 30-s video spots on safety belt use, featuring young males in the leading roles, were developed and shown to the personnel. Also presented were demonstrations with a crash simulator in which volunteers experienced actual *g* forces of low-speed crashes, and talks by traffic safety experts.

The fine for not using a safety belt when riding in the front seat of a vehicle while at these bases varied from a warning to a fine of Dfl. 35 (about $ 17) to Dfl. 65 (about $ 30). During the second month of the campaign, drivers and their front-seat passengers at their respective bases were punished (i.e., either warned or fined) or rewarded, dependent upon intervention type. The military police departments involved were requested to report the number of hours they spent on surveillance and how many fines they recorded during the enforcement campaign.

3.2.3 Experimental conditions

Three levels of enforcement were implemented: (a) minimal (4 hr of surveillance), (b) moderate (16 hr), and (c) intensive (32 hr) during the last four weeks of the campaign. Level of enforcement was factorially combined with type of publicity, resulting in six experimental conditions, distributed over eight military bases.

Four additional bases served as experimental groups to investigate effects of incentive programs. On these four incentive-bases, no enforcement was implemented. Two bases had a contest or group-dependent incentive during the campaign; Dfl. 5,000 (about $ 2,500) in cash could be won by the whole group of personnel at the base showing the highest belt use rate at the end of the campaign. The prize was to be spent on a party or other activity, in such a way that all personnel could benefit from the prize. On two other bases, an individual incentive program was implemented during the second month of the campaign, whereby lottery tickets were distributed to drivers and front-seat passengers who were observed using their safety belts. Incentive rates were varied as follows: on one base one prize was drawn every week, whereas on the other base four prizes were drawn every week. Prizes included money coupons, photocameras, portable cassette tape players, and compact disc players (see Table 3.1 on p. 46 for an overview of the experimental design).

3.3 Results

A total of 21,671 obtrusive observations were taken, of which 10,888 occurred during baseline, 5,642 immediately following the campaign, and 5,141 during follow-up observations. A total of 17,072 unobtrusive observations were made, of which 7,486 occurred during baseline, 3,495 immediately following the campaign, and 2,922 during follow-up.

Figure 3.1 depicts the percentages of drivers using a shoulder belt during each condition. In most, but not all, cases, the obtrusive observations differed significantly from the unobtrusive observations. Overall mean safety belt use was 71.9% during baseline obtrusive observations, 83.1% directly following the campaign, and 86.2% during follow-up. Overall mean belt use during unobtrusive observations was 65.4% during baseline, 72.9% directly following the campaign, and 76.1% during follow-up.

3.3.1 Log-linear analyses

A log-linear analysis (using the CATMOD-procedure provided by SAS Institute, 1985) was conducted on data collected unobtrusively with belt use as the dependent variable and both base (1 through 12) and experimental phase (1 through 3) as independent variables. This analysis revealed significant main effects for both base, $\chi^2(df = 11) = 215.9;\ p < 0.001$, and experimental phase, $\chi^2(df = 2) = 55.0;\ p < 0.001$. The interaction between base and experimental phase was also significant, $\chi^2(df = 21) = 156.9;\ p < 0.001$. At seven bases, observed belt use increased during the period of study (bases 1, 4, 5, 6, 7, 11, and 12 in Figure 3.1), and at five bases no increase was found (bases 2, 3, 8, 9, and 10 in Figure 3.1). These results reflect no systematic variation with treatment condition, thereby complicating an interpretation of the results.

When the experimental conditions were divided into two groups (enforcement vs. incentive programs) a log-linear analysis revealed significant main effects of program type, $\chi^2(df = 1) = 25.4;\ p < 0.001$, and experimental phase, $\chi^2(df = 2) = 88.8;\ p < 0.001$, but no significant interaction between program type and experimental phase, $\chi^2(df = 2) = 1.1;\ p > 0.50$.

This suggests that both enforcement and incentive programs had the same overall effect. Enforcement and incentive programs showed mean increases from 67% and 62% during baseline to 75% and 69% directly after the campaign, and to 78% and 76% during follow-up, respectively.

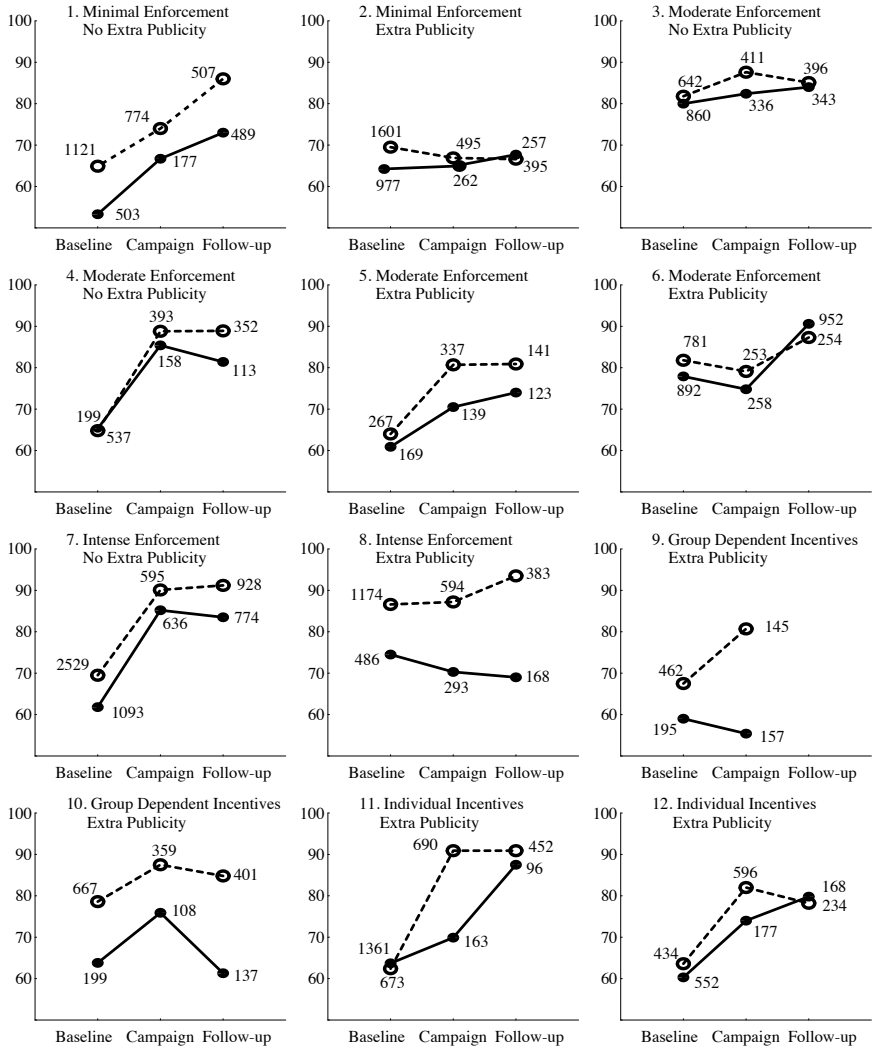

Figure 3.1: Percentages of drivers using a safety belt for each condition and experimental phase. Open circles connected by dotted lines represent results of obtrusive observations; solid circles represent results of unobtrusive observations. The numbers associated with each data point indicate the number of observations.

When the bases were divided into two groups, those with extra publicity (EP) and those without extra publicity (P) during the campaign, a main effect of amount of publicity, $\chi^2(df = 1) = 12.7$; $p < 0.001$, was found, as well as a significant interaction between amount of publicity and experimental phase, $\chi^2(df = 2) = 83.5$; $p < 0.001$. Interventions of P and EP, respectively, showed observed safety belt use rates of 65% and 69% during baseline, of 80% and 70% directly after the campaignand of 80% and 75% during follow-up, respectively. Thus, an unexpected inverse relationship between type of publicity and increase in belt use was observed.

3.3.2 Age and personnel groups

To investigate the effects of age and personnel group, additional log-linear analyses were performed on the data obtained with the obtrusive observations. A significant main effect of age was obtained, $\chi^2(df = 1) = 177.9$; $p < 0.001$, as well as an interaction between age and experimental phase, $\chi^2(df = 2) = 29.1$; $p < 0.001$. Also, a main effect of personnel group was found, $\chi^2(df = 1) = 168.1$; $p < 0.001$, but the interaction between personnel group and experimental phase was not significant, $\chi^2(df = 2) = 3.2$; $p > 0.05$. Figure 3.2 illustrates these results. Because the second-order interaction between base, age and personnel group was not significant, $\chi^2(df = 11) = 16.0$; $p > 0.10$, the observations were averaged over all 12 sites. Observed values for civilian personnel were left out of these comparisons because of the very small sample size for this group.

Figure 3.2 shows that drafted personnel younger than 25 years old had an average safety belt use of 58% during baseline, increasing to 77% directly after the campaign, and to 83% during follow-up. Enlisted personnel younger than 25 had an observed belt use of 68% during baseline, 85% directly after the campaign, and 87% during follow up. Drafted personnel 25 years old or older had an observed belt use of 80% during baseline, 79% directly after the campaign, and 82% during follow-up. Enlisted personnel of this age group showed 84% belt use during baseline, 89% directly after the campaign, and 90% during follow-up. Thus, the observed increases in belt use throughout the entire period of study were almost entirely due to a belt use increase of 37.7% above baseline by drivers under 25 years of age (from 61% to 84%). Drivers of 25 years or older showed an increase of only 6.0%. It should be noted, however, that their baseline level was much higher (an increase from 84% to 89%).

Log-linear analyses revealed no main effect of observation day, $\chi^2(df = 1) = 3.48$; $p > 0.05$, entering versus departing traffic, $\chi^2(df = 1) = 0.22$; $p > 0.50$; and no interaction between these variables, $\chi^2(df = 1) = 0.02$; $p > 0.50$.

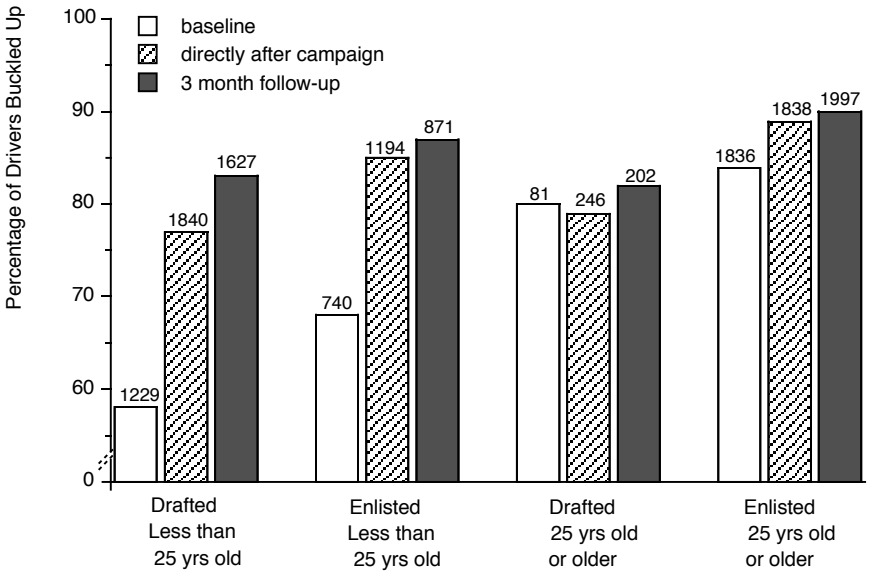

Figure 3.2: Observed safety belt use for drivers between 18 and 24 years-of-age vs. 25 years-of-age or older, for drafted and enlisted personnel. The number associated with each bar indicates the number of observations.

3.3.3 Actual police effort

A post-hoc analysis of actual military police effort found that it varied from base to base, ranging from 16 to 72 hr of surveillance and from 4 to 100 fines registered (see Table 3.1 on p. 46). Treatment allocations were not reflected in actual reported hours nor number of fines. However, the number of fines per hour had a correlation of 0.80 (Pearson r, $p < 0.05$) with observed increases in belt use.

3.4 Discussion

The effects found were almost entirely attributable to a dramatic belt use increase of 37.7% for young drivers between 18 and 25 years old. Because the campaign was aimed at this group of drivers in particular, this result fulfilled an important campaign objective. Drivers 25 years old and older showed only a 6.0% increase

over baseline, but because their baseline level was already above the 70% mean belt use in the Netherlands, this was not really surprising.

Quite unexpectedly, different baseline belt use rates for drafted personnel and enlisted personnel were found (especially for drivers under 25 years old). Previous research has found safety belt promotion campaigns to have differential effects on blue-collar versus white-collar personnel (Geller & Bigelow, 1984); our findings might reflect similar population differences.

No systematic effects of enforcement level (as originally anticipated) upon belt use were found. It was difficult to understand why the same amount of enforcement led to an enormous increase in safety belt use on one base but had no effect at another. However, the actual effort of the police (in terms of number of fines per hour) was related to these differences: The more police effort, the higher the observed increase in belt use. This suggests that enforcement indeed enhances safety belt use substantially, *if* the enforcement is actually carried out. It is realized that a linear regression model based on only seven observations (because of missing values for one base; see Table 3.1 on p. 46) is hardly convincing. In addition, 'number of fines per hr' might not be the most suitable measure of police effort. However, no other information relevant to amount of enforcement was available.

The awareness of the campaign was rather disappointing. No effect of extra publicity on observed belt use was found when publicity (P) and extra publicity (EP) treatments were compared. Counter to expectations, the P condition showed greater impact than the EP condition. Results from a written survey (Hagenzieker, 1990) revealed that only 40% of the respondents reported to have noticed an increase in publicity during the campaign. The results from the questionnaire also showed that respondents from the EP treatment conditions did not report to have noticed the publicity more often than respondents from the other conditions. In addition, the military police who distributed the publicity materials reported an insufficient supply of handouts. Therefore, it can tentatively be concluded that extra publicity might have been effective if the personnel had in fact been exposed to the extra materials.

An important finding in the present study was that incentive programs are capable of enhancing safety belt use beyond high initial baseline belt use rates. This holds especially for the individual incentive programs (INC1 and INC4), which showed medium-term increases. Because the differences in effects between the two conditions (one prize per week vs. four prizes per week) were rather small and not statistically significant, it can not be determined which strategy is more effective. The results of the group-dependent incentive (GR) program showed at best only a short-term effect. Because a whole group won the prize, including the nonusers of safety belts, this condition might also be considered a noncontingent

reward program; noncontingent rewards are known to have less impact than contingent rewards, such as those applied in the individual incentive treatments (e.g., Geller et al., 1982b). Therefore, in accordance with previous findings, this contingent versus noncontingent distinction may explain the differences found for the individual versus group incentive treatments.

On average (i.e., across all conditions) safety belt use had not decreased three months after withdrawal. In contrast, on most bases it had actually increased even further compared to observations directly after the campaign. These results were counter to expectations. An optimistic interpretation of the results obtained would be that the campaign was successful in establishing long-term maintenance of the target behavior. A probably more realistic interpretation is that the police, stimulated by the campaign, continued or even increased enforcement efforts after the formal end of the campaign.

In a recent study, Kalsher et al. (1989) investigated the relative impact of incentive and enforcement ('disincentive') programs on two US naval bases. Their study is comparable to the present one in many respects. Kalsher et al. found that, at one navy base, the use of safety belts increased by 10 percentage points during a 4-week incentive program which was similar to the results of the individual incentive programs we employed; at another naval base a 24 percentage point increase was observed by Kalsher et al. during the 3-week enforcement phase. These investigators attributed the greater impact of the disincentive program almost entirely to a dramatic increase in belt use when vehicles entered the gates. (In the present study, no effect of entering versus departing traffic on observed belt use was found.) With regard to departing vehicles and follow-up, the impact of the two intervention approaches (i.e., incentives vs. disincentives) was equivalent in the Kalsher et al. study. As in the previous study, in the present study the average impact of enforcement and incentive interventions was the same, and was also of about the same size as reported by Kalsher et al. We found a medium-term increase of 11 percentage points on the enforcement treatment bases and an increase of 14 percentage points on the incentive treatment bases. Moreover, the baseline levels were comparable (i.e., about 60% buckled up in both studies).

Finally, a number of weaknesses in the present study should be mentioned. First, the study lacked a nonintervention control group. Unfortunately, when conducting field studies it is often not possible to include a control group. The field study involving the experimental treatments could only take place when implemented in an already organized nationwide campaign at all military bases in the Netherlands during the same period, or not at all. The lack of control groups complicated the interpretation of the results, because only comparisons of effects relative to each other could be established.

Second, the target behavior of all experimental groups should ideally be comparable before the interventions. However, because safety belt use was never studied previously in a military setting in the Netherlands, no matching on the basis of baseline belt use was possible.

A third complicating factor was that the two observation methods yielded different results. In general, belt use was higher during the obtrusive than during the unobtrusive observations. The presence of military police during the obtrusive observations may have prompted (or activated) general "careful" behavior by drivers, including the use of a safety belt.

Chapter 4

Drivers' opinions of enforcement and incentive strategies to promote safety belt use*

During a nationwide campaign to promote safety belt use among military personnel, a field study was conducted at 12 military bases in the Netherlands. Enforcement and incentive programs were varied among military bases. A written survey was administered to the personnel of these bases. The survey contained items concerning reported belt use, motivation to use a safety belt, attitudes toward legislation, public information, enforcement and incentive strategies, and awareness of the campaign. The results showed that enforcement was clearly a better accepted countermeasure than rewarding drivers for the use of safety belts. Respondents from bases exposed to the incentive treatment tend to have relatively more positive opinions of incentives than those exposed to the enforcement treatment. Two independent dimensions were present in the response patterns: one representing opinions of enforcement and the other of incentives/rewards.

*Reprinted from Journal of Safety Research, 23(4), 199–206 (Hagenzieker, 1992b), with permission from Elsevier Science.

4.1 Introduction

Although most drivers recognize that vehicle safety belts are effective in reducing or preventing driver injuries, many do not use safety belts. In the Netherlands, a safety belt mandate was enacted in 1975; national belt use rates increased from 20% in 1974 to 50% in 1975 and to 70–75% in 1990 (Verhoef, 1991). However, young male drivers use their safety belts less often than other groups. Thus, legislation alone is not sufficient to achieve (nearly) universal use of safety belts. Young male drivers are also frequently involved in accidents (Twisk, 1990); thus safety belt use is especially critical for this group. Therefore, a nationwide campaign was designed to promote safety belt use among military personnel, a population made up primarily of young males.

During this two-month campaign, which began in October 1988, a field study was conducted at 12 military bases in the Netherlands. During the campaign the personnel of eight bases were exposed to an enforcement campaign, the personnel of the other four bases to an incentive campaign. The first month was used to announce the campaign. For those bases assigned to incentive treatments, special brochures were disseminated that outlined the intervention condition. Personnel at those bases assigned to enforcement treatments received brochures containing information about penalties for not buckling up. Other publicity materials included stickers, playing cards, and posters with the special campaign logo printed on them. In addition, local and military newspapers, and radio stations provided some media coverage about the campaign. During the second month of the campaign, drivers and their front seat passengers were punished (i.e., either warned or fined) or rewarded, dependent on intervention type. On the four incentive bases no enforcement was implemented. Rewards were either group dependent (i.e., two bases had a contest that would be won by the whole group at the base showing the highest belt use rate at the end of the campaign) or individual (i.e., at two bases once a week prizes were drawn ranging from money coupons to compact disc players on the basis of lottery tickets that were distributed to drivers and front seat passengers who were observed using their safety belts). The military police departments involved kept a record of hours they spent on surveillance and how many fines they recorded during the enforcement campaign; they were also involved in distributing the lottery tickets during the incentive campaign.

Observations of safety belt use among servicemen in their personal vehicles were conducted before the campaign, immediately following the campaign, and three months later. Enforcement and incentive campaigns had the same overall effect. Enforcement and incentive programs showed mean increases from 67% and 62% during baseline to 75% and 69% directly after the campaign, and to 78% and 76% during follow-up, respectively. To a large degree the effects were due to a 37.7%

increase among young drivers, less than 25 years-of-age (Hagenzieker, 1991). The results of this study confirm previous findings that both enforcement (e.g., Grant et al., 1991; Gundy, 1988) and incentive strategies (e.g., Elman & Killebrew, 1978; Geller, 1988; Kalsher et al., 1989) can increase safety belt use rates substantially.

In general, drivers' attitudes toward legislation and enforcement – in countries with a safety belt use mandate – are moderately positive (e.g., Gundy, 1988; Milosevic & Pajevic, 1988). Little is known about drivers' opinions of incentive programs to stimulate safety belt use. A survey conducted in Canada showed that incentives were rated by the public among the favorite countermeasures to increase traffic safety (Wilde et al., 1975). To investigate drivers' opinions of enforcement and incentive programs and their relations with treatment types, an observational study (see Hagenzieker, 1991) as well as a written survey was carried out among the personnel of the bases involved in the various types of campaigns. The questionnaire contained items concerning reported belt use, motivation to use or not use a safety belt, attitudes toward enforcement and incentive strategies, and awareness of the campaign.

4.2 Method

4.2.1 Participants and setting

A total of 3,000 questionnaires were distributed among the personnel of the 12 military bases directly after the end of the campaign, in December 1988. The eight enforcement bases received a total of 2,000 questionnaires, the four incentive bases a total of 1,000. The number of questionnaires received were dependent upon the size of the respective bases. The staff of the bases was requested to hand out the survey to personnel between 18- and 25 years-of-age, who had been on the base during the campaign. They were instructed to have the questionnaires completed by the respondents, who remained anonymous, individually (and not collectively) within two weeks, and return them in the enclosed postage-free envelopes.

4.2.2 Materials

The questionnaire contained a total of 57 items, of which a subset of 31 is reported in this article (the 26 remaining ones were more detailed items on campaign content). They were distributed over five main topics: (a) reported safety belt use – in general and as compared to two months ago, 2 items; (b) awareness of the campaign – in general and with respect to the enforcement and incentive programs, 4

items; (c) motives to use a safety belt, 8 items; (d) attitudes toward legislation, public information, enforcement and incentives, 12 items; (e) background variables – such as age, base, and whether at the base during the campaign, 5 items.

4.2.3 Data analysis

Data analysis involved several steps. First, respondents who never drove a car to and from the base and those who had not been on the base during the campaign were excluded from further analysis. Then, frequencies for all variables were calculated. Differences in responses between treatment types were calculated. And finally, to explore the underlying structure of answers related to enforcement and incentives, a Principal Components Analysis was carried out

4.3 Results

4.3.1 Response and background variables

Out of the 3,000 distributed questionnaires a total of 1,341 (45%) was returned; 1,032 from the enforcement bases (52%), and 309 from the incentive bases (31%). The discrepancy between the response percentages was due to the fact that one of the incentive bases did not return any questionnaires. Of these respondents, 14% never drove a car; they were therefore excluded from further analysis. Most of the respondents (95%) reported to have been on the base during the campaign. Eventually, 1,097 of the returned questionnaires could be used for further data analysis.

Over half (55%) of these 1,097 respondents indicated to drive a car to and from the base on a daily basis. Only 16 respondents were female, which was to be expected at military bases. The majority (78%) of the respondents were 25 years old or younger; 61% of the respondents were drafted personnel, 39% were enlisted personnel. None of the background variables had significant relations with all other variables reported below.

4.3.2 Reported safety belt use

Most of the 1,097 respondents (64%) reported to use their safety belt "always," 18% said "often," 8% "sometimes," 6% "seldom," and 4% "never." When asked

whether they used the safety belt now (i.e. at the end of the campaign), more often, equally often or less often than two months ago, a majority of 83% answered to wear their safety belt "equally often" and 15% answered "more often."

The combined results of these two items showed that those who had reported to use their belt "often" or "sometimes" also reported to use their belt "more often" now, 32% and 35%, respectively. Of those respondents who had reported "seldom" or "never" to use a safety belt, only 11% indicated to use their belt "more often" now. No relation between these variables and treatment type was found.

4.3.3 Awareness of the campaign

Although 79% indicated to have read, heard or noticed otherwise "something about the usefulness of using safety belts" during the past two months, only 40% was of the opinion that this publicity had increased over the past two months; 55% said the amount of publicity on this topic had remained the same, and 5% even said it had decreased. Respondents from 'incentive bases' reported significantly more often that publicity had increased (45%) than respondents from 'enforcement bases' (35%; $\chi^2 = 12.1, df = 3, p < 0.01$).

About two-thirds of the respondents (68%) indicated to have noticed police enforcement of safety belt use at the base during the past two months. Respondents from enforcement bases reported to have noticed this enforcement significantly more often (70%) than those from incentive bases (59%; $\chi^2 = 10.4, df = 2$, $p < 0.01$).

One-third of the respondents reported to have noticed "something" about rewarding drivers for using a safety belt at military bases. This was usually the case at bases exposed to the incentive treatments (75%); 20% of the respondents from bases exposed to the enforcement treatment were also aware of drivers being rewarded for using a safety belt ($\chi^2 = 227.9, df = 2, p < 0.001$), but they usually noticed this from reports in the newspaper, whereas respondents from 'incentive-bases' more often actually experienced the rewarding.

4.3.4 Motives to use a safety belt

Table 4.1 gives an overview of the differences between treatment types for variables concerning motives and attitudes. In general, the most important motives to use a safety belt were: (a) because of its protective properties in case of crashes (indicated by 76% of the respondents); (b) habituation (60%); and (c) because it

Table 4.1: Summary of differences between treatment types for variables concerning motives and attitudes. Variables are listed by their original item numbers (English translations of the Dutch items). Responses were on a five-point scale; categories 1 and 2 are shown combined – agree (completely) – as well as categories 4 and 5 – disagree (completely). The last column shows χ^2 values ($df = 5$); * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

		TREATMENT TYPE									
		ENFORCEMENT (N=849)				INCENTIVES (N=248)					
		response category				response category					
VARIABLE		1+2	3	4+5	missing	1+2	3	4+5	missing		
MOTIVES		%	%	%	%	%	%	%	%	χ^2	
"I use a safety belt because":											
(5c)	it is compulsory	59	13	13	15	54	13	12	21	*NS*	
(5d)	I don't want trouble with the police	44	15	26	15	34	17	28	21	11.74	*
(5g)	of rewards	7	9	66	17	12	10	58	21	13.32	*
"I would use a belt more often when":											
(6e)	the police would check more often	21	16	41	22	20	12	41	27	*NS*	
(6f)	the fine was higher	21	16	39	24	22	12	38	28	*NS*	
(6j)	I could get rewarded	12	15	49	24	21	9	43	27	22.83	***
ATTITUDES											
(22a)	"the police have more important things to do than checking on belt use"	63	22	13	2	55	21	20	3	12.17	*
(22b)	"it is a just cause that the police check on safety belt use"	54	26	17	3	60	23	12	4	*NS*	
(23a)	"it is childish that the police check on belt use"	27	18	52	4	33	15	49	4	13.55	*
(23b)	"it is good when the police reward for belt use"	57	20	20	3	49	19	28	4	15.16	**
(26a)	"rewarding is an original way to stimulate belt use"	48	16	33	3	60	14	23	3	13.55	*
(26b)	"rewarding safety belt use is a waste of money"	47	20	29	4	33	19	41	6	20.31	**
(26c)	"the police show their right side by rewarding belt use"	32	25	38	5	42	23	29	6	15.67	**
(26d)	"it is exaggerated to reward belt use since it is compulsory"	66	17	14	3	52	16	28	5	30.61	***

is compulsory (58%) (item 5c, see Table 4.1). Equally often respondents indicated that more enforcement or higher fines would cause them to use their safety belt more often (both about 20%; 6e and 6f). No differences between treatment types were found for these variables.

Three items showed significant differences between treatment types: while 44% of therespondents from enforcement bases indicated to use a belt because they "don't want trouble with the police," 34% of respondents from incentive bases indicated this to be an important motive (5d). On the other hand, respondents from incentive bases reported more often that rewards are an important motive to use a safety belt than those from enforcement bases (5g and 6j; see Table 4.1).

4.3.5 Attitudes toward legislation, public information, enforcement and incentives

In general, over three-quarters (78%) of the respondents completely agreed with the statement that "public information on the use of safety belts is good," and 67% completely disagreed with the statement that this information is "a waste of time and effort." With regard to legislation 83% approved of the safety belt use mandate. However, still 35% of the respondents were of the opinion that one has to decide for him- or herself whether to use a belt or not. No differences between treatment types were found for these variables.

Averaged over treatment types, 56% were of the opinion that the police have better things to do than checking on belt use (22a), while 61% thought it is a just cause that the police check on belt use (22b); 28% thought fining is "childish" (23a), while 55% agreed with the fining of non-users (23b). About half (51%) of the respondents agreed with the statement that rewarding is "an original way to stimulate safety belt use" (26a). On the other hand, 43% agreed with the statement that "rewarding is a waste of money" (26b). About one-third (34%) thought that by rewarding the police show their "right side" (26c), whereas 63% felt it is exaggerated to reward for using a safety belt – since it is compulsory (26d). All variables concerning attitudes toward incentives showed significant differences for treatment type (see Table 4.1). In general, respondents from bases exposed to the incentive treatment had relatively more positive attitudes toward incentives than those exposed to the enforcement treatment (26a–d). However, this trend does not show up in the answers for item 23b, which shows more positive attitudes toward incentives for respondents from the enforcement condition. Only minor, and not systematic, differences between treatment types with regard to enforcement were found (22a, 23a; see Table 4.1 on p. 60).

4.3.6 Principal components analysis

To explore underlying response-structures with regard to enforcement and incentives, the 14 items concerned with motives (6) and attitudes (8) were entered in a PCA. The eigenvalues of the solution for the first and second dimension were 3.23 and 2.64, respectively, meaning that both dimensions together describe 42% of the variance (the eigenvalue for a third dimension was considerably less, namely 1.76). Figure 4.1 shows the component scores of the 14 variables (indicated by their original item number in the questionnaire) as well as the object scores (representing the respondents). Variables 22, 23 and 26 concerned 'attitude'-items; variables 5 and 6 'motives.' In general, the attitude items loaded higher on both dimensions than the items concerning motives. Items concerning rewards loaded generally high on the first dimension, while items concerning enforcement loaded generally high on the second dimension.

4.4 Discussion

Inspection of the results of the principal components analysis revealed that the first dimension can be interpreted as representing the respondents' opinions of incentives/rewards; objects located on the left of Figure 4.1 represent respondents with negative opinions of incentives, while objects on the right represent positive opinions of incentives. The second dimension can be interpreted as representing opinions of enforcement: the lower part of Figure 4.1 shows respondents in favor of enforcement, while the upper part shows those who have objections to enforcement. Item 23b ("it is good when the police reward for belt use") loaded relatively high on the second dimension together with items concerning enforcement. No logical explanation seems to be available for this finding, although it could be possible that respondents associated this item with 'enforcement' since it is phrased in terms of 'the police.' Objects on the extreme right of the figure represent relatively more often respondents from incentive bases (see also Table 4.1 on p. 60). Since opinions of enforcement did not vary systematically between treatment types but opinions of incentives did (see Table 4.1), one could infer from these results that regardless of their attitude toward enforcement (positive or negative) respondents can adjust their opinion of incentives dependent on treatment type experienced during the campaign.

All combinations of opinions were present, but the combination 'pro enforcement – contra incentives' was the most frequent one for both respondents from enforcement and incentive bases (about 30%; cf. lower-left corner in Figure 4.1). About 15% were against both enforcement and rewards (cf. upper-left corner); about

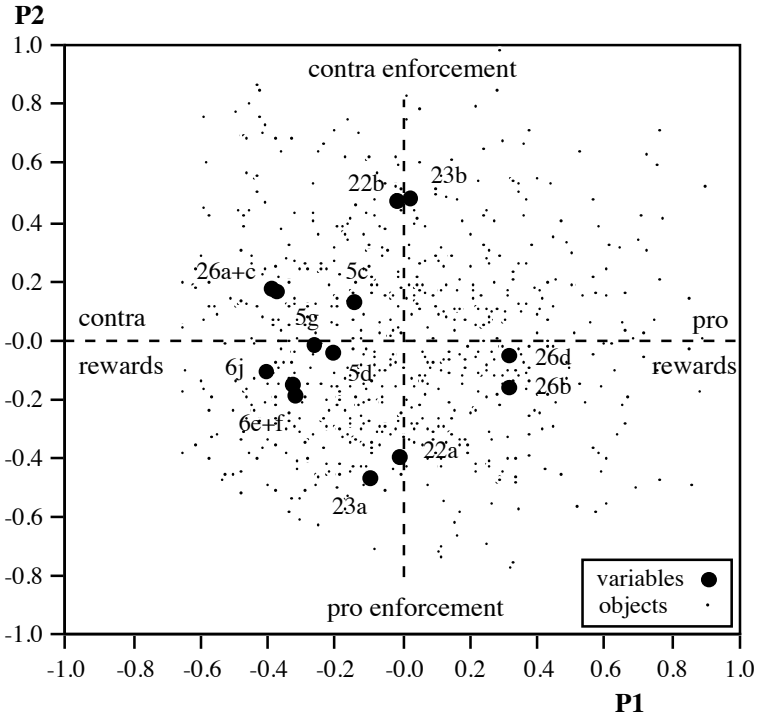

Figure 4.1: Results of two-dimensional (P1, P2) principal components analysis, showing component scores of 14 variables as well as the object scores (N=728; eigenvalues 3.23, 2.64; variance 0.23, 0.19). The first dimension can be interpreted as representing opinions of incentives/rewards, the second as opinions of enforcement.

10% of the respondents were pro both enforcement and rewards (cf. lower-right corner), and about 10% were for rewarding but against enforcement (cf. upper-right corner). The remaining 35% of the respondents had more neutral opinions.

In general, enforcement is clearly better accepted than rewarding drivers for 'good behavior' by both respondents from enforcement and incentive bases. This can be derived from the clear positive opinions of enforcement as opposed to those of incentive programs. The positive attitudes toward enforcement confirm earlier findings (e.g., Gundy, 1988; Milosevic & Pajevic, 1988). Contrary to results obtained elsewhere (Wilde et al., 1975), this study shows that incentives are not rated among the most popular countermeasures, at least so among (young) Dutch military personnel. However, these not-so-positive attitudes toward incentives did not

result in a smaller effect on actual safety belt use than the more accepted enforcement treatment. As Hagenzieker (1991) showed, both treatments had the same overall effect: an increase of 11%-points with baseline levels of about 65%.

About two third (64%) of the respondents indicated to use their safety belt "always;" and 15% of the respondents said they used a safety belt more often as compared to before the campaign. Although the resemblance between actual and reported belt use figures is remarkable (see also Streff & Wagenaar, 1989), it cannot be concluded from these data that reported usage truly reflects actual usage. The fact is that the actual belt use of the individual respondents was not known, only overall belt use rates per base, which can be considered a shortcoming of this study.

One would expect that respondents exposed to the incentive treatment would have noticed enforcement the least, as opposed to those in the enforcement conditions, which was indeed found. However, still about 60% of the respondents from the incentive bases had noticed enforcement during the campaign, although no enforcement should have been applied at all. The mere presence of the police could have caused the association with 'enforcement' (but no fines were delivered at the incentive treatment bases).

In general, treatment types were clearly visible in the respondents' response patterns. Especially the variables concerning 'incentives' yielded different results for the two treatment types. Since the respondents of the two treatment types did not differ systematically in any of the background variables nor in their opinions of legislation and police enforcement, it is very unlikely that they had different opinions with regard to incentives from the start. However, whether the campaign caused drivers to change their behavior (see Hagenzieker, 1991) and consequently their attitudes toward incentives, or vice versa, remains unclear (see also, McGuire, 1985).

Characteristic of the items concerned with motives for using a safety belt is the large number of missing data; over 20% (see Table 4.1). This was also encountered in earlier research using comparable questions (Gundy, 1986). Apparently, respondents find this type of questions hard to answer, and skip them consequently.

It should be mentioned that the general awareness of the campaign was disappointing. While a comparable (enforcement) campaign in the Netherlands was able to achieve about 60% of the public to notice an increase in publicity (Gundy, 1988), only 40% of the respondents in the present study noticed this. This could be explained by a shortage of publicity materials. When asked, the police who distributed the materials indeed reported an insufficient supply of handouts.

The findings of the present study would lead one to assume that combining both enforcement and incentive strategies during a campaign could result in larger effects than either of them alone, since the various measures are favored by different groups of people; this was in fact reported recently (Mortimer et al., 1990). However, it should be realized that 'being in favor' of a certain measure does not automatically result in a behavior change. This is illustrated by those respondents in the present survey who reported to use their safety belt "seldom" or "never;" only 11% of them indicated to use their belt "more often" two months after the campaign as compared to before the campaign.

Finally, it should be realized that this study covers a special population of young males, surveyed in a special environment following special campaign types to stimulate belt use. Therefore, generalization of the present results beyond these special circumstances must be viewed with caution. Further research, for example, with other sets of participants, should reveal whether the same patterns of results also apply to more general parts of the population.

Chapter 5

Effects of incentives on safety belt use: A meta-analysis*

The effects of campaigns using tangible incentives (rewards) to promote safety belt usage have been evaluated by means of a meta-analytic approach. Two coders extracted a total number of 136 short-term and 114 long-term effect sizes and coded many other variables from 34 journal articles and research reports. The results show a mean short-term increase in use rates of 20.6 percentage points; the mean long-term effect was 13.7 percentage points. Large scale studies report smaller effect sizes than small scale studies; when studies were weighted by the (estimated) number of observations, the weighted mean effect sizes were 12.0 and 9.6 percentage points for the short- and long-term, respectively. The main factors that influence the magnitude of the reported short-term effect of the programs were the initial baseline rate (which was highly correlated with the presence or absence of a safety belt usage law), the type of population involved, whether incentives were delivered immediately or delayed, and whether incentives were based on group or individual behaviour. Together these four variables accounted for 64% of the variance. Other variables, such as the duration of the intervention, the probability of receiving a reward, and the value of the reward were not related to the short-term effect sizes. The relationship between moderating variables and long-term effects was less clear.

*Reprinted from Accident Analysis & Prevention, Vol.29(6), 759–777 (Hagenzieker et al., 1997), with permission from Elsevier Science.

5.1 Introduction

In the past 15–20 years a large number of incentive programs to stimulate safety belt usage have been implemented and evaluated, most of them in the USA. Their initial application appears to have been inspired by the absence of mandatory safety belt use laws in many states at the time – excluding enforcement activities as a possibility to influence wearing rates – and by learning theory (Skinner, 1953) and behaviour modification techniques (e.g., Kazdin, 1984). More recent incentive programs have been conducted to investigate their effect as an alternative for or in addition to enforcement strategies to increase safety belt use.

A problematic aspect in reviewing the results of individual studies is the poor systematic variation in campaign characteristics and settings, or a subset of these aspects within studies. This problem can, partly, be overcome by choosing a meta-analytic approach instead of a more 'traditional', narrative literature review to analyse the reported results and other variables of incentive programs to promote safety belt use; meta-analysis is considered to be better able to integrate research findings across studies. According to Glass et al. (1981 , p.21) "The essential character of meta-analysis is that it is the statistical analysis of the summary findings of many empirical studies. [...] it is not a technique; rather it is a perspective that uses many techniques of measurement and statistical analysis". Advantages of meta-analysis over narrative reviews that are often expressed include the feature of being systematic – and therefore more explicit, more exhaustive, and more replicable – and that conclusions are often more specific (Mullen, 1989; Rosenthal, 1991, 1995; see also Hedges & Olkin, 1985; Hunter & Schmidt, 1990; however, see, e.g., Lepper, 1995 for a more critical approach to meta-analysis as a theoretical tool). To date, the method of meta-analysis has hardly been used in reviewing research findings in the area of evaluating road safety campaigns aimed at changing road user behaviour. The reviews by Moore (1989), Elliott (1993) and Johnston et al. (1994) are the only three examples of meta-analyses in this area known to the authors (although Johnston et al. do not use the term 'meta-analysis' in their article).

Moore (1989) conducted a meta-analysis of mass media campaigns designed to change automobile occupant restraint behaviour. He distinguished three types of campaign: selective traffic enforcement campaigns (STEPs), message-only, and incentive campaigns. However, only one incentive campaign was included in this meta-analysis (Campbell et al., 1984; Gemming et al., 1984; Hunter et al., 1986 these publications were combined into one entry), because one of the criteria for inclusion in his meta-analysis was that the campaign made use of incentives in an electronic mass media campaign and most incentive campaigns did not have an electronic mass media component. Elliott (1993) has performed a meta-analysis

in order to determine the effects of road safety mass media campaigns, including 39 restraint use (safety belts and child restraints) campaigns. His analysis showed that campaigns to stimulate restraint use result in an average relative increase of ca 50% over baseline. However, only one study using incentives (Elman & Killebrew, 1978) was included in Elliott's analysis. Johnston et al. (1994) compared the effectiveness of five types of behavioural safety belt programs and concluded that incentive programs, as well as law and enforcement practices produced the largest increases in safety belt use during intervention (medians of 17.6 and 14.4 percentage points, respectively). Length of intervention was not significantly related to usage rates. The authors did not further differentiate between specific contents of the various program types in their analysis.

Other (narrative) reviews (e.g., Geller, 1984b; Geller et al., 1987; Thyer & Geller, 1990) have reported that incentive interventions typically result in a doubling of baseline safety belt use during or directly after the end of the intervention. However, these results have been obtained with relatively low baseline levels of safety belt use. Later studies on incentive programs, conducted under conditions where mandatory seat belt laws were in effect – therefore with much higher baseline usage rates – have also shown substantial increases in belt use (e.g., Hagenzieker, 1991; Kalsher et al., 1989). In these latter studies, effect sizes of incentive treatments have been reported as being comparable to those of enforcement treatments (ca 10 percentage points with baseline safety belt usage rates ca 50–60%). Although the specific contents of the incentive programs often vary, a first inspection of the findings indicates that all types of intervention seem to give similar results. For instance, rewards have been given directly or indirectly, immediate or delayed, they could be contingent or noncontingent upon actual safety belt use, the probability, type and value of the reward could vary, they have been based on group or individual behaviour etc. Also the type of population in which the program was implemented varies from work settings to schools and universities, or even large community settings. Long-term effects are usually reported as being smaller than short-term effects, but the follow-up belt usage rates are generally higher than the initial baseline levels. However, more specific (quantitative) effect sizes are not available. Besides the general conclusion that various kinds of incentive programs seem to work equally well, it is not known which characteristics of the incentive programs are more effective than others, and to what extent, and in which settings such programs are most effective (cf. Peters, 1991).

The aim of the present study is to determine the short- and long-term effects of incentive programs to stimulate safety belt use. Furthermore, it has been attempted to locate moderating variables that are related to the magnitude of these effects. To this end a large number of study-characteristics, which were selected on the basis of previous empirical findings or theoretical considerations, have been coded for a sample of studies. This review is confined to studies that have used behavioural

observations of safety belt use, and to programs that included incentives that could be expressed in terms of some 'material value' (such as cash, free meals, prizes that could be won in a contest or lottery etc.) as opposed to incentives such as praise, approval or social attention.

5.2 Method

5.2.1 Sample of studies

A literature search using IRRD (International Road Research Documentation) and PsycINFO databases was conducted to identify studies on the topic of incentive programs to promote safe driving behaviour. The databases were searched on the following key words: 'incentive' or 'reward', in combination with 'safety', and 'driving/driver behaviour'. There was no constraint on the year of publication, nor on the publication form. The studies were gathered until the end of January 1994, with the intention of including all earlier, relevant studies. Using the same key words, Current Contents and the SWOV-library have been consulted weekly for the last five years. Reference lists from articles were also reviewed, and all research on the topic that was known to the authors was included as well. A total of 160 articles and research reports were found. Then the type of study was checked (review, theoretical article, or research report/article), and only those publications (n=91) reporting on (experimental) research findings were selected for the analysis.

Studies were excluded from the analysis:

(1) when the incentive/reward was not given to individuals, but, for example, organizations or policy makers;

(2) when the incentives/rewards were nonmaterial, for instance praise or social attention;

(3) when in addition to incentives, enforcement was part of the same intervention program;

(4) when no behavioural measures were presented (e.g., questionnaires);

(5) when the article was a review.

After the publications that met up to one or more of the exclusion criteria were deselected, the remaining number of publications was 54. Of those remaining publications, 40 had the subject of seat belt programs (as opposed to other behaviours

or outcomes). The last step was to exclude documents reporting on the same experiments (e.g., Geller & Hahn, 1984; Gemming et al., 1984; Hunter et al., 1991b; Stutts et al., 1984). A complete list of excluded documents is available from the first author; see Appendix for the final selection of 34 articles. A literature search performed in June 1996 suggests that since January 1994 no studies have been published that would have been included in the present meta-analysis). In summary, the inclusion criteria for the studies to be analysed in this meta-analysis were that the studies used incentives with material value based on the observed safety belt use of individuals. Twenty-three variables, which were selected on the basis of previous empirical findings or theoretical considerations, were used to describe the characteristics of the studies. The variables used to describe the studies, grouped according to the type of information they give, were:

Background variables: (1) Year of publication; (2) publication form; (3) country (and state in case of USA) in which the program was carried into effect; (4) the presence or absence of a safety belt use law; (5) type of population; (6) population size; (7) whether the population had previously participated in another incentive program; (8) psychological theory related to incentives/rewards as referred to by the author(s).

Research characteristics: (a) Research design, (b) observation method (obtrusive or unobtrusive observations), (c) length of the follow-up period in days, (d) the number of observations in each of the three observation periods – baseline, intervention and follow-up.

Characteristics of the incentive program: (1) Duration of the intervention period recorded in days; (2) whether the receiving of rewards depended on the behaviour of the group or on the behaviour of each individual; (3) whether the delivery of the reward(s) was immediate or delayed; (4) whether the rewarding was contingent or noncontingent upon the actual usage of a safety belt; (5) the number of different kinds of reward given; (6) type of incentive/reward (categorized in the same way as Geller, 1982: exchangeable token, immediate valuable, promotional item, chance to win a contest, work related privilege); (7) value of the incentive/reward (in US-dollars); (8) the probability for those who use safety belts to receive the incentive; (9) the expected value of the reward (i.e. value $\times$ probability).

Information about the impact of the program and effect sizes: Belt usage during the observation periods was recorded. In most cases percentages of belt

use were reported for three phases: (a) before the incentive program (baseline period), (b) during (or directly after), and (c) some time after the program ended (follow-up period). Seat belt usage was expressed as a percentage of seat belt users among an observed sample of car passengers. Effects of incentive programs were calculated from changes in seat belt usage. Meta-analysis can be based on correlations (Hunter & Schmidt, 1990), or on effect sizes or differences between means (e.g., Hedges & Olkin, 1985). Here, the latter approach was considered the most appropriate, because the original research usually reports effects in terms of before-and-after values. Two effect measures were distinguished, a short-term effect $(b-a)$, and a long-term effect $(c-a)$ (cf. d'; see Rosenthal, 1991). When, if at all, a control group was included in the study, the effect observed in the control group was subtracted from the effect in the experimental group. (The common effect size estimate used in meta-analyses of experimental studies incorporates the information of a control group $((X_E - X_C)/s_C)$ (see also, for example, Glass's Δ, Cohen's d, Hedges's g; Rosenthal, 1991; Hunter & Schmidt, 1990; Hedges & Olkin, 1985). Because this information was rarely available, this measure could not be used in the present analysis. The alternative effect sizes for studies with control groups were used side by side with the 'ordinary' effect sizes for studies without control conditions.

5.2.2 Coding procedure

Two of the authors independently coded each of the studies on each of the variables (an example of the coding sheet that was used is available from the authors). When the required characteristic was not reported in a study, coders were instructed to attempt to deduce the information needed (e.g., from graphs) or to make a knowledge-based estimate (e.g., the value of a named reward, for example when the reward was a 'T-shirt' or 'compact disc player' of which no monetary value was specified). It was expected to be difficult for the coders to determine whether or not studies as reported within the same document should be considered as separate treatments – in particular when more than one intervention took place. (Many publications reported on more than one study or intervention.) It was therefore decided to keep the information of both coders in the database (instead of compromising between the two codings) and to add the variable 'coder' as a factor in the further analyses.

5.2.3 Analysis steps

All steps in the analysis were performed both on the data of the two coders separately as well as on the combined data. First, basic statistics were calculated.

Next, a homogeneity analysis using the program HOMALS (Gifi, 1981) was performed to investigate whether relationships appear to exist between the variables of interest. In particular, we were interested in exploring the relationships between the effect sizes and potentially moderating variables to be used in further analyses. HOMALS attempts to position studies with similar characteristics close to each other and relatively 'dissimilar studies' farther away from each other. In this way the studies are divided into various 'homogeneous' groups. The location of each study in the plot is decomposed into contributions of each variable. In this manner the apparent effect of each variable can be compared. Most importantly, the effect variables can be compared with the other variables, resulting in a selection of apparently relevant variables for further analyses. After these first orientations, general linear model (GLM) type analyses were performed (SAS Institute, 1990) to identify the variables that could explain differences between the effect sizes – dependent variables – of different kinds of incentive programs. A GLM-analysis is a general form of regression and variance analysis with the use of linear modelling that can handle both continuous and categorical variables.

Furthermore, the following checks on the general sensitivity of the obtained GLM-outcomes have been performed.

Transforming the effect measure A separate analysis was carried out using a power of the effect size measure instead of the effect size itself. Its results were compared to the original results. The reason for making this change originated from the supposed influence the base rate has on the effect size that can be achieved. It is assumed to be easier to increase the belt use rate by 40 percentage points starting off from a rate of 20% seat belt usage compared to an initial rate of 60%; a power transformation (e.g., $b^2 - a^2$ instead of $b - a$) emphasizes effects starting at higher baseline rates.

Sensitivity of the analysis to relative importance of the observations It can be assumed that larger sample sizes lead to more reliable measures of belt use. However, because the precise number of observations was often not reported, effect sizes could not be weighed by sample sizes (besides, it did not seem appropriate to weigh studies with, for example, 100,000 observations 100 times as compared to studies with a sample size of 1000). Therefore, estimates of sample sizes in four categories were used to check their influence on the results, using various weight distributions. The weights were given as follows: effect sizes based on: (1) $<$ 250 observations: class 1; (2) 250–1000 observations: class 2; (3) –10,000: class 3;(4) $>$10,000: class 4. Apart from using the weights 1–4, also other values have been used that satisfied the ordinal relationship.

Stability The stability of the GLM-results themselves was assessed using a bootstrap analysis (see, for example, Efron & Tibshirani, 1993). This was done by sampling from both the studies and coders, assessing not only the unreliability of the model but also the unreliability caused by the differences between the coders.

5.3 Results

As expected, it appeared difficult for the coders to determine whether or not studies as reported within the same publication should be considered as separate treatments, which resulted in unequal numbers of identified studies (61 by coder 1 and 78 by coder 2). A total of 139 studies (including duplicates) were coded by the two coders – an average of two studies per publication. Table 5.1 shows a summary of study characteristics.

5.3.1 Background variables

Most studies were published in journals (81%) and were conducted in the USA (94%). The studies were published between 1978 and 1992 (median 1986). In the majority of cases (83%) no mandatory safety belt use law was in effect at the time of study. Programs are typically implemented in companies (31%), communities (27%) or schools (34%); at high schools and universities the programs are usually aimed at students and staff, at elementary schools at both children and their parents. The (estimated) size of the population that was involved in the incentive campaign varied from less than 50 persons (e.g., various substudies in Geller, 1989b) to over 100,000 (Mortimer, 1991). About one fifth of the populations had previously been involved in another incentive program. The difference between the two coders in the number of identified treatments is reflected in the scoring of this variable: coder 2 indicated more frequently that the incentive program was preceded by another one in the same article. In the majority of cases (60%) no reference was made to psychological theories related to the effects of rewards/incentives; learning theory and behaviour modification techniques (e.g., Skinner, 1953; Bandura, 1977, 1986; Kazdin, 1984) were referred to in 16% of the studies, 19% mentioned motivational theories (e.g., Deci, 1978; Deci & Ryan, 1985; Festinger, 1957), and 4% referred to utility, expectancy or cost-benefit theories (e.g., Tversky & Kahneman, 1974; Atkinson, 1964).

Table 5.1: Summary of study characteristics

Variable and class	Total Value[†]		Coder 1 Value		Coder 2 Value	
Background variables:						
Median publication year	1986		1986		1986	
Publication form						
Journal	112	(80.6%)	48	(78.7%)	64	(82.1%)
Report	27	(19.4%)	13	(21.3%)	14	(17.9%)
Country and/or state of study						
USA Virginia	59	(42.4%)	24	(39.3%)	35	(44.9%)
USA rest	72	(51.8%)	33	(54.1%)	39	(50.0%)
Netherlands	8	(5.8%)	4	(6.6%)	4	(5.1%)
Mandatory safety belt use law						
Yes	24	(17.3%)	9	(14.8%)	15	(19.2%)
No	115	(82.7%)	52	(85.2%)	63	(80.8%)
Type of population						
Community	37	(26.6%)	18	(29.5%)	19	(24.4%)
Company (incl. military)	43	(30.9%)	22	(36.1%)	21	(26.9%)
University or high school	19	(13.7%)	7	(11.5%)	12	(15.4%)
Elementary school	28	(20.1%)	11	(18.0%)	17	(21.8%)
Else	12	(8.6%)	3	(4.9%)	9	(11.5%)
(Estimated) Population size	13,665	(1,100)	18,243	(1,075)	10,078	(1,167)
Study was preceded by other incentive program in the same population.						
Yes, within same article	17	(12.2%)	4	(6.6%)	13	(16.7%)
Yes, in other article	14	(10.1%)	12	(19.7%)	2	(2.6%)
No	108	(77.7%)	45	(73.8%)	63	(80.8%)
Theory referred to						
Learning theory/ behavior modification	22	(15.9%)	4	(6.6%)	18	(23.1%)
Motivational theories	27	(19.4%)	7	(11.5%)	20	(25.6%)
Utility/cost-benefit theories	6	(4.3%)	5	(8.2%)	1	(1.3%)
No theory mentioned	84	(60.4%)	45	(73.8%)	39	(50.0%)
Research characteristics:						
Research design						
AB	18	(12.9%)	5	(8.2%)	13	(16.7%)
AB + C	3	(2.2%)	–	–	3	(3.8%)
AB(D)A	94	(67.6%)	43	(70.5%)	51	(65.4%)
ABA(D) + C	17	(12.2%)	98	(13.1%)	9	(11.5%)
Else	7	(5.0%)	5	(8.2%)	2	(2.6%)
Observation method						
Obtrusive	67	(48.2%)	37	(60.7%)	30	(38.5%)
Unobtrusive	55	(39.6%)	14	(23.0%)	41	(52.6%)
Obtrusive and unobtrusive	15	(10.8%)	10	(16.4%)	5	(6.4%)
Unknown	2	(1.4%)	–	–	2	(2.6%)

[†]For categorical variables, numbers represent frequencies and percentages (in parentheses) in each class; for continuous variables means and medians (in parentheses) are given; belt use is expressed in percentage wearing rate.

Table 5.1: Summary of study characteristics (continued).

Variable and class	Total Value		Coder 1 Value		Coder 2 Value	
Follow-up period in days	74.5	(42)	84.6	(56)	65.7	(26)
(Estim.)N. of baseline obs.	6,185	(733)	9,076	(680)	3,999	(788)
(Estim.)N. of intervention obs.	6,279	(781)	9,618	(822)	3,884	(771)
(Estim.)N. of follow-up obs.	6,579	(633)	9,645	(548)	4,097	(690)
Characteristics of incentive program:						
Duration of intervention in days	44.6	(28.0)	44.0	(28.0)	45.0	(28.0)
Group or individual reward						
Group	20	(14.4%)	12	(19.7%)	8	(10.3%)
Individual	113	(81.3%)	46	(75.4%)	67	(85.9%)
Both group and individual	6	(4.3%)	3	(4.9%)	3	(3.8%)
Delivery of rewards						
Immediate	52	(37.4%)	20	(32.8%)	32	(41.0%)
Delayed	56	(40.3%)	23	(37.7%)	33	(42.3%)
Both immedaite and delayed	31	(22.3%)	18	(29.5%)	13	(16.7%)
Contingency of rewards						
Contingent	109	(78.4%)	55	(90.2%)	54	(69.2%)
Noncontingent	30	(21.6%)	6	(9.8%)	24	(30.8%)
Type of reward‡						
Work related privilege	1	(0.5%)	1	(1.2%)	–	–
Exchangeable token	60	(33.0%)	22	(25.6%)	38	(40.0%)
Immediate valuable/ promotional item	57	(31.3%)	25	(29.1%)	32	(33.3%)
Chance to win contest/lottery	64	(35.2%)	38	(44.2%)	26	(27.1%)
Number of different rewards	1.9	(1)	2.5	(2)	1.4	(1)
Value of reward	$282	($9)	$306	($11)	$265	($4.8)
Probability of receiving a reward	0.28	(0.19)	0.40	(0.25)	0.19	(0.15)
Expected value	73.3	(0.75)	145.0	(0.99)	17.4	(0.47)
Safety belt use:						
Belt use during baseline	27.7%	(23.1%)	26.8%	(23.0%)	28.4%	(23.8%)
Belt use immediately after interv.	48.8%	(45.7%)	47.5%	(45.4%)	49.7%	(46.9%)
Belt use after follow-up period	43.1%	(39.3%)	40.5%	(38.9%)	45.4%	(39.6%)

5.3.2 Research characteristics

The typical research design applied for evaluating the behavioural effects of incentive programs to promote safety belt use is the ABA design, which involves observations of safety belt use before intervention, during or immediately after intervention, and some time after withdrawal of the program (67%). Control groups

‡More than one category could be valid; therefore, n does not sum up to 139

(C) are not often involved in the research design (AB+C:2.2%; AB(D)A+C:12.2%; three studies in category 'else': 2.1%; a total of 16.5% with control group). Sometimes more than one intervention (D) was involved in the design (ABDA: 4.3%; ABDA + C: 1.4%). The number of observations per phase varied from less than 50 (e.g., various substudies in Lehman & Geller, 1990) to over 50,000 (Mortimer et al., 1990). The observation method could be either obtrusive (48%), unobtrusive (40%), or both obtrusive and unobtrusive (11%). The follow-up period varied from 5 (Geller et al., 1989) to 546 days (Nimmer & Geller, 1988).

5.3.3 Campaign characteristics

The duration of the intervention varied from one day (e.g., Elman & Killebrew, 1978) to 365 days (Robertson, 1984), with a median value of 28 days. The value of the reward varied from US$ 1 (in many campaigns) to several thousands of dollars (e.g., Hagenzieker, 1991; Horne & Terry, 1983a; Robertson, 1984: $2,500–$10,000), and the probability to receive a reward varied from almost 1 (e.g., Geller, 1982; Hagenzieker, 1989) to almost zero (e.g., Foss, 1989; Hunter et al., 1991a). Rewards were most often based on individual behaviour (81%) as opposed to the behaviour of a whole group of people (14%) or to both group and individual behaviour (4%), and were in most cases contingent upon safety belt use (78%). An example of a noncontingent reward (22%) is when all individuals who signed a pledge card in which they promise to use a safety belt could get a prize, and usage was not checked. Seat belt wearers were rewarded at the spot ('immediately': 37%) or at a later stage ('delayed', for example, in case of a lottery: 40%), or a combination of both (22%). Rewards/incentives were exchangeable tokens such as cash or meal coupons (33%), immediate valuables or promotional items such as stickers and T-shirts (31%), or a chance to win a contest or lottery (35%). Usually one type of reward was offered, sometimes more than one type (mean 1.9).

5.3.4 Differences between coders

To compare the results of the two coders, association measures of matched records from the two coders (related to the same treatments identified), were calculated for the categorical (Cramer's V, CV) and continuous variables (correlations). Correspondence between coders was best for the variables indicating belt usage in various phases (Pearson r between 0.94 and 0.97), duration of intervention ($r = 0.95$), value of reward/incentive ($r = 0.96$), duration of the follow-up period ($r = 0.86$), and the background variables country (and state in case of USA) of study, and year of publication (CV=0.94). These characteristics were all clearly reported (association measures are not 1.0 because the number of entries of coder 1 and 2

were not the same, and sometimes belt use for a given phase had to be inferred from graphs). The other characteristics were often not explicitly mentioned in the articles, and the coders had to infer the desired information from the text. Therefore, association measures of most variables were between 0.60 and 0.70. Worst correspondence was found for the variables: number of rewards ($r = 0.38$), type of reward ($r = 0.40$), observation method (CV=0.50), and preceding studies (CV=0.58). Coder 1 systematically identified a greater *number* of different *types of reward* in the studies than coder 2, which resulted in low association measures for these two variables. The observation method (obtrusive and/or unobtrusive) was rarely mentioned explicitly, and it appeared difficult to deduce this from the procedures-section of the original papers. Coder 2 more often indicated that the study was preceded by another incentive program than coder 1 (see also *background characteristics*).

As a second check of the comparability of the two codings, the factor 'coder' was added as a variable in the GLM analyses (see *Impact of moderating variables*) to investigate whether any difference between the coders was systematic, causing these differences to influence the results.

5.3.5 Effect sizes

Short-term effect sizes A total of 136 short-term effect sizes were calculated (in three cases no short-term effect size could be calculated), 58 by coder 1 and 78 by coder 2. The minimum calculated effect was slightly negative (Robertson, 1984), the maximum effect was over 60 percentage points (Roberts & Fanurik, 1986). Figure 5.1a shows a histogram of the short-term effect sizes. Figure 5.2a shows the weighted means (and 95% confidence limits) of the short-term effect size for the various weight classes. The (unweighted) mean short-term effect was 20.6 percentage points ($\pm$2.4; median 19.5). Pearson r between coder 1 and 2 was 0.88. The means for coder 1 and 2 were 19.9 and 21.1 percentage points, respectively.

The number of observations on which the percentages of belt use were based varied enormously between studies. It appears that the largest studies on average report the smallest effect sizes. The mean of weight class 1 (effect sizes based on < 250 observations) equals 24.0 percentage points ($\pm$5.0); the mean of weight class 4 ($> 10,000$ observations) 6.9 percentage points ($\pm$3.3). The maximum overall weighted mean is reached when all studies have equal weights: 20.6 percentage points. The weighted mean approaches a minimum of 6.9 percentage points when it is based mainly on the largest studies. When studies were weighted by the (estimated) number of observations, the mean short-term effect size was 12.0 percentage points ($\pm$1.9).

Long-term effect sizes Not all studies included follow-up observations; a total of 116 long-term effect sizes were calculated, 53 by coder 1 and 63 by coder 2. The mean long-term effect was 13.7 percentage points (±1.8; median 12.0; Pearson *r* between the coders was 0.70). The means for coder 1 and 2 were 13.3 and 14.0 percentage points, respectively. The minimum calculated long-term effects were slightly negative (Geller et al., 1982b, 1989; Hunter et al., 1991a); the maximum calculated effects were over 30 percentage points (Geller, 1989b; Lehman & Geller, 1990). Figure 5.1b shows a histogram of the long-term effect sizes. The mean of weight class 1 equals 18.2 percentage points (±3.9); the mean of weight class 4 7.8 percentage points (±2.1; see Figure 5.2b). The maximum overall weighted mean is reached when all studies have equal weights: 13.7 percentage points. The weighted mean approaches a minimum of 7.8 percentage points when it is based mainly on the largest studies. When studies were weighted by the (estimated) number of observations, the mean long-term effect size was 9.6 percentage points (±1.2).

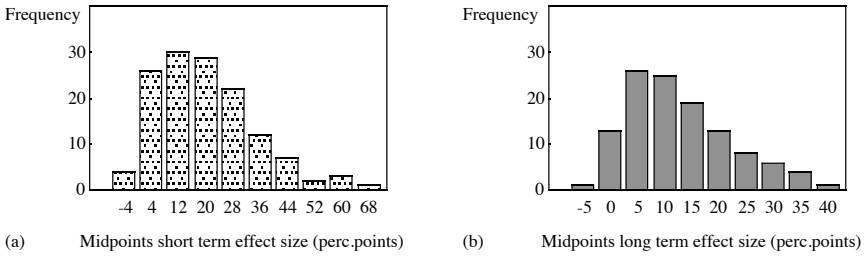

Figure 5.1: Histogram of (a) short-term and (b) long-term effect sizes.

5.3.6 Publication bias

Funnelplots (see, for example, Mullen, 1989) were produced to inspect whether certain types of results (e.g., zero or very small effects) are not or sparsely present in the published documents. The idea of the funnelplot is based on the operation of the law of large numbers, which states that the larger the sample size, the more probable it is that the sample mean is a good estimate of the population mean. On the horizontal axis of a funnelplot the effect sizes are shown, on the vertical axis the sample sizes on which the effect sizes were based. According to the law of large numbers, there will be less variation in the effect sizes of the large sample

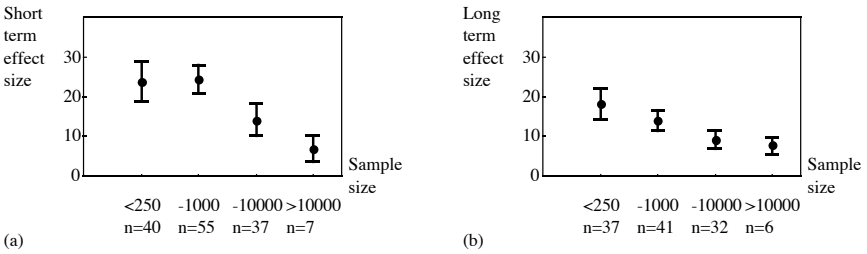

Figure 5.2: Weighted means and 95% confidence limits of the (a) short-term and (b) long-term effect sizes according to weight class. Class 1, effect sizes based on < 200 observations; class 2, –1000 observations; class 3, –10,000 observations; class 4, $> 10,000$ observations.

sizes, and more variation in the effect sizes of smaller sample sizes. If there is no publication bias, the plot should resemble this situation, resulting in a distribution that takes the shape of an inverted funnel. Figure 5.3 shows that the distribution of data points is somewhat skewed in favour of large short-term effect sizes, and that only few large studies are present (which have relatively small effect sizes). The funnelplot of the long-term effect sizes looked very similar and is not shown here.

Another way to look at the problem of publication bias is to check whether studies published in journals report larger effect sizes than those in unpublished research reports (cf. Lipsey & Wilson, 1993); this was not the case for the present data (mean short-term effect sizes were 21.2 and 18.0 percentage points (standard error of the mean 1.4 and 2.8) for journal articles and reports, respectively; mean long-term effect sizes were 14.4 and 9.7 percentage points (standard error of the mean 1.0 and 1.5); all nonsignificant differences (Tukey-Kramer)).

5.3.7 Exploratory analyses

To explore underlying structures with regard to effect sizes and background characteristics all variables were entered in a HOMALS-analysis. After removing variables with very low discrimination values, 17 variables remained in the analysis. The eigenvalues of the solution for the first and second dimension were 0.26 and 0.24, respectively, meaning that both dimensions together describe 50% of the variance. Figure 5.4 shows the category quantifications for these variables. The solutions for the data of the two coders were similar (but somewhat rotated as compared) to the one obtained for the combined data.

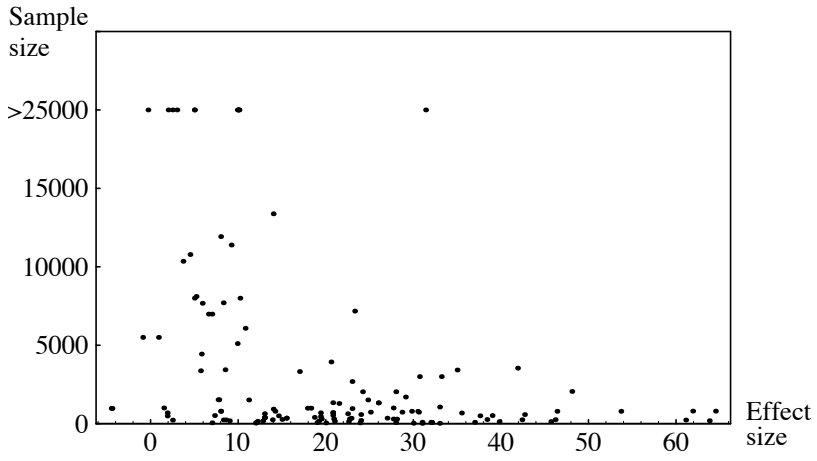

Figure 5.3: Funnelplot of short-term effect sizes by sample size (see text for explanation).

Figure 5.4 shows that the effect sizes (subdivided into a number of classes) primarily load on the first dimension (horizontal axis); on the left side of the figure small effect sizes are seen, to the right the larger effect sizes. The variables 'type of population', 'sample size' (small, moderate, large), and 'promotional item' (yes, no) also load primarily on the first dimension, indicating that particular categories of these variables often co-occur with particular effect sizes. Variables that load primarily on the second dimension are: 'country and/or state of study', 'group and/or individual reward', 'number of rewards', 'presence of a safety belt use law' (yes, no), and the 'presence of a control group' in the research design.

Some studies were clustered at the outer edge of the solution, fairly isolated from the rest of the studies (note that datapoints representing individual studies are not shown in Figure 5.4). These appeared to be studies by the same author, or substudies in the same publication. These studies have systematically used the same unique (combination of) categories in their study (see, for example, Roberts et al., 1990; Roberts & Fanurik, 1986; Roberts et al., 1988); these studies were all located at the bottom of the solution, near the categories 'group and individual rewards', 'elementary school', 'value $2–10', etc. Consequently, generalizations of the findings of this type of studies to other situations are not possible.

The HOMALS-solution suggests the presence of moderator variables that influence the magnitude of the effect sizes (for example, the variables 'type of popu-

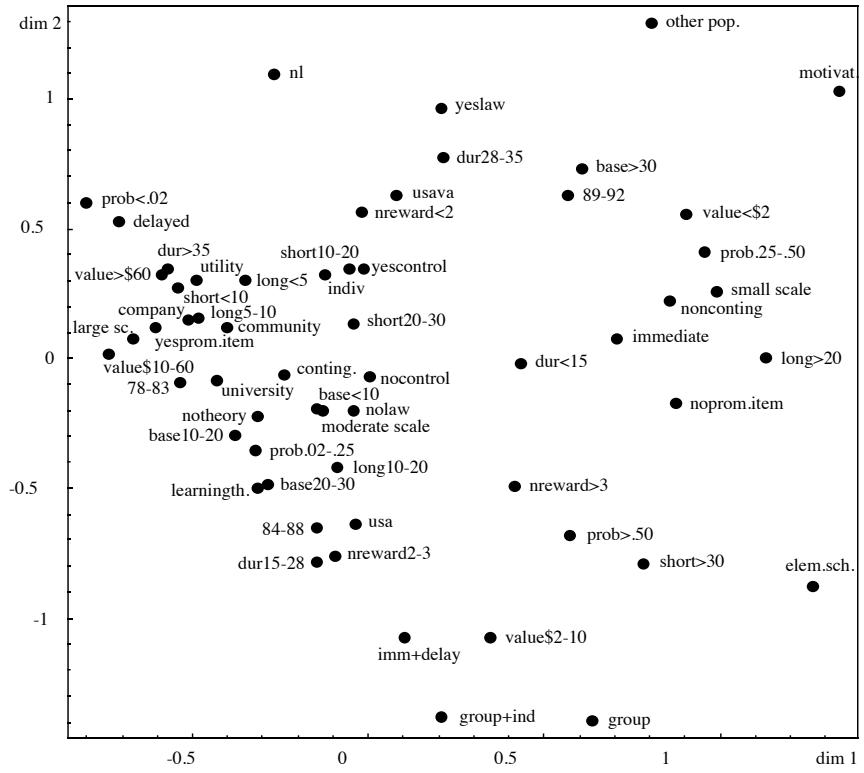

Figure 5.4: Results of the two-dimensional HOMALS analysis, showing category quantifications of 17 variables: duration of incentive program in days (dur <15;15–28;28–35;>35); probability to receive a reward (prob <0.02; 0.02–0.25; 0.25–0.5;>0.5); number of different types of reward in program (nreward <2;2–3;>3); short-term effect size in percentage points (short <10;10–20;20–30;>30); long-term effect size in percentage points (long <5; 5–10;10–20;>20); baseline safety belt use rate in percentages (base <10;10–20;20–30;>30); mandatory safety belt use law (yeslaw, nolaw); theory referred to [motiv(ational), utility, learning th., no theory]; estimated value of reward in $ (value <$2; $2–10; $10–60; >$60); control group in research design (yescontrol, nocontrol); population involved [company, community, elem(entary) sch(ool), (high school or) university, other pop(ulation)]; immediate and/or delayed delivery of rewatds (immediate; delayed; imm + delay); contingency of rewards [conting(ent); nonconting(ent)]; publication year (1978–1983; 1984–1988; 1989–1992); country [USA–Virginia; USA(–other states); NL]; size of study (small scale – weight class 1; moderate scale – class 2; large scale – class 3 and 4); group and/or individual rewards [group; indiv(idual); group + ind].

lation', 'sample size', and 'promotional item' that also load primarily on the first dimension); it also shows that there is a lot of confounding in the characteristics of the studies (see, for example, clusters of studies using the same combination of characteristics) and that certain combinations of characteristics are simply not present in the dataset. For example, studies published in more recent years (1989–1992) tend to have relatively high baseline levels ($> 30\%$ belt use); large short-term effects (> 30 percentage points) tend to go together with elementary schools as population type, and high probabilities (> 0.50) to receive a reward; delayed delivery of rewards tends to be connected with low probabilities of receiving one (< 0.02), and with relatively high values of the rewards ($> \$\ 60$).

5.3.8 Impact of moderating variables

First, all variables were entered in a GLM-analysis as single main factors to investigate how much variance of the short- and long-term effect sizes each of them could account for. Weighted, unweighted, and adjusted effect sizes were subsequently used as dependent variable.

For some variables, tests of fit of single variable models did not show significant F-values for any of the *short-term* effect measures; these were: 'coder', 'value of the reward', 'probability to receive a reward', 'number of different rewards', the reward types 'exchangeable token' and 'chance to win a contest', 'duration of intervention', whether the study was 'preceded by another incentive program', 'observation method', and publication year and form. Other variables always had significant F-values, regardless of the type of short-term effect measure; these were: 'type of population' – accounting for 35–40% of the variance, depending on the type of dependent variable –, 'group and/or individual reward' (10–20% of the variance), 'immediate and/or delayed delivery of rewards' (10–15% of the variance), and the presence of 'promotional items or immediate valuables' (5–10% of the variance). Note that two of those variables were also brought out in the homogeneity analysis.

The single main factor that accounted for most of the variance of the *long-term* effect sizes was the short-term effect, explaining 40–46% of the variance for the unadjusted effect sizes (Pearson $r = 0.62$); for adjusted effect measures the influence was reduced to 10–13%. In addition, the following variables always had significant F-values, regardless of the long-term effect measure: 'type of population' (accounting for 12–24% of the variance, depending on the type of dependent variable), 'immediate and/or delayed delivery of rewards' (14–24% of the variance), 'theory referred to' (12–24%), 'contingency of rewards' (10–15% of the variance), the presence of 'promotional items or immediate valuables' (8–14%), and the 'probability to receive a reward' (5–13% of the variance).

Finally, tests of fit for some single variable models were significant with part of the effect measures (weighted/unweighted; adjusted/unadjusted) as explanatory variables. A summary of the results of the GLM-analyses with single variables is presented in Table 5.2.

The differences of the means for each category of the variables were tested by means of post hoc comparisons (Tukey-Kramer). Table 5.3 shows an overview of the means and standard errors of the mean per category of a number of variables; significant differences are indicated by asterisks. Incentive programs implemented in elementary schools had much larger effect sizes than those in the other types of population. Group based rewards led to larger effect sizes than individual rewards, the effect being more pronounced for the short-term effect measures; delayed delivery of rewards led to smaller effect sizes than immediate delivery of rewards, and the presence of promotional items or immediate valuables as rewards led to larger effect sizes than when this type of rewards was absent. These were more pronounced for the unweighted and adjusted effect measures, meaning that promotional items were present relatively more often in small scale studies, and with relatively high baseline levels.

Short-term effect sizes were larger when no mandatory safety belt use law was in effect during the incentive campaign; under these circumstances baseline belt use was relatively low: 22.5% belt use as compared to 52.7% with mandatory belt use law (see also Table 5.2: a main significant effect of 'usage rate during baseline'). Low baseline levels were connected with relatively high short-term effect sizes (Pearson $r = -0.25$; $p < 0.01$). The presence or absence of a mandatory safety belt use law did not have any effect on the (unadjusted) mean long-term effect size, nor when adjusted short-term effect sizes were used in the GLM-analyses. When the research design included a control group both short- and long-term effect sizes were smaller than when no control group was present. The main effect of 'control group' was only significant when weighted effect sizes were used in the GLM-analyses; it appears that in particular large scale studies (with higher weights) had included control groups. Counter to expectations, campaigns with contingent rewards showed smaller effect sizes than those with noncontingent rewards; the difference was more pronounced for long-term effect measures and for adjusted measures (see Table 5.2 on page 85), which implies that noncontingent rewards were relatively more often applied in settings with high baseline levels. For long-term effect sizes, and when adjusted short-term effect sizes were used in the GLM-analyses, the main factor 'theory referred to' was significant; larger effect sizes were reported when motivational theories were referred to as compared to 'no theory referred to'. In addition, for adjusted long-term effects main effects of 'publication year' (more recent studies had high baseline levels and relatively large long-term effect sizes), and 'country and/or state of study' were significant (Netherlands differed from USA because of high baseline levels and relatively

Table 5.2: Summary of GLM-analyses with single variables and short-term effect size (a) and long-term effect size (b) as dependent variables; degrees of freedom ($df_{1,2}$), F-values, significance level (p), and proportion of variance accounted for (r^2) are shown for weighted, unweighted and adjusted effect size measures. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

	Weighted				Unweighted			Adjusted, weighted			Adjusted, unweighted		
Variable	$df_{1,2}$	F	p	r^2	F	p	r^2	F	p	r^2	F	p	r^2
(a) *Short-term effect*													
Mandatory safety belt use law	1,135	20.10	***	0.131	11.00	*	0.076	2.77	NS	0.020	0.32	NS	0.002
Type of population	4,135	17.65	***	0.352	17.30	**	0.347	22.67	***	0.409	22.97	***	0.412
Control group	1,135	11.14	**	0.077	4.94	NS	0.036	5.88	*	0.042	2.87	NS	0.021
Group and/or individual reward	2,135	15.54	***	0.191	17.24	**	0.207	11.20	***	0.144	9.87	***	0.129
Immediate and/or delayed delivery	2,135	10.48	***	0.137	14.95	**	0.185	10.37	***	0.134	12.79	***	0.161
Contingency of rewards	1,135	7.19	**	0.051	3.76	NS	0.027	20.05	***	0.130	16.15	***	0.107
Theory referred to	3,135	2.67	NS	0.057	1.71	NS	0.038	4.87	**	0.099	6.47	**	0.128
Promotional item	1,135	4.58	*	0.033	8.23	*	0.058	9.48	**	0.066	17.35	***	0.115
Usage rate during baseline	1,135	12.71	***	0.087	8.81	*	0.062	1.03	NS	0.008	4.04	*	0.046
(b) *Long-term effect*													
Publication year	2,115	0.01	NS	0.000	1.67	NS	0.029	6.33	**	0.101	12.01	***	0.175
Country and/or state of study	2,115	1.10	NS	0.019	0.53	NS	0.009	9.24	***	0.141	5.84	**	0.094
Mandatory safety belt use law	1,115	1.26	NS	0.011	0.00	NS	0.000	5.78	*	0.048	11.10	**	0.089
Type of population	4,115	8.80	***	0.243	8.73	***	0.241	3.91	**	0.123	4.60	**	0.142
Control group	1,115	7.21	**	0.060	6.09	*	0.051	1.53	NS	0.013	1.92	NS	0.017
Group and/or individual reward	2,115	4.83	**	0.079	3.21	*	0.054	0.31	NS	0.005	0.29	NS	0.005
Immediate and/or delayed delivery	2,115	12.67	***	0.184	17.24	***	0.235	8.83	***	0.135	15.17	***	0.212
Contingency of rewards	1,115	12.81	***	0.102	11.57	***	0.093	15.38	***	0.119	19.62	***	0.147
Theory referred to	3,115	4.87	**	0.116	5.07	**	0.121	9.32	***	0.200	12.05	***	0.244
Promotional item	1,115	17.90	***	0.137	30.52	***	0.213	9.25	**	0.075	17.34	***	0.132
Exchangeable token	1,115	0.62	NS	0.005	2.91	NS	0.025	6.99	**	0.058	9.68	**	0.078
Probability of receiving reward	1,115	15.99	***	0.124	16.56	***	0.128	6.21	*	0.052	5.24	*	0.044
Usage rate during baseline	1,115	0.34	NS	0.003	1.20	NS	0.011	43.58	***	0.277	49.40	***	0.302
Short-term effect size	1,115	93.66	***	0.460	73.67	***	0.401	16.07	***	0.127	12.72	***	0.104

Table 5.3: Unweighted and weighted means and standard error of the mean per category of each variable for short- and long-term effects (asterisks indicate significant differences between means; $p < 0.05$, Tukey-Kramer test).

		Short term effect				Long-term effect			
		Unweighted		Weight 1–4		Unweighted		Weight 1–4	
Variable	Category	Mean	Stderr	Mean	Stderr	Mean	Stderr	Mean	Stderr
Mandatory safety belt use law	yes	12.0 *	2.2	8.9 *	1.8	13.6	2.3	10.3	1.9
	no	22.5 *	1.4	21.4 *	1.3	13.6	1.0	12.5	0.9
Type of population	community	13.4	1.7	11.3	1.6	9.1	1.5	7.8	1.1
	company	20.0	1.8	19.9	1.8	11.9	1.1	12.2	1.0
	university	17.4	2.7	16.9	2.8	9.7	2.3	9.1	2.0
	elem.school	36.5 *	2.9	37.4 *	2.9	21.6 *	2.0	20.6 *	1.9
	other	13.3	3.1	13.1	3.0	15.5	3.5	12.0	3.2
Control group	no	21.8 *	1.4	20.6 *	1.3	14.6 *	1.0	13.0 *	0.9
	yes	14.1 *	2.8	10.3 *	2.3	8.9 *	1.8	7.8 *	1.4
Group and/or individual reward	individual	17.5 *	1.1	16.0 *	1.1	12.4 *	1.0	10.8	0.9
	group	32.5 * *	4.0	30.4 * *	3.9	17.9 *	1.9	16.8	1.9
	group and individual	38.8 *	8.3	35.7 *	7.5	16.8	2.2	16.7	2.1
Immediate and/or delayed delivery	immediate	22.2 *	1.6	20.7	1.7	19.0 *	1.8	16.4 *	1.6
	delayed	14.2 *	1.3	13.4 *	1.2	8.7 * *	0.9	8.4 * *	0.8
	immediate and delayed	30.2 *	3.7	26.0 *	3.6	16.4 *	1.7	14.9 *	1.6
Contingency of rewards	contingent	19.4	1.3	17.2	1.3	11.9	0.8	10.7	0.7
	noncontingent	25.2	3.0	25.6	3.1	18.7	2.4	17.5	2.2
Promotional item	no	17.8 *	1.3	16.8	1.3	9.6	0.9	9.6 *	0.8
	yes	24.9 *	2.3	22.2	2.3	18.5	1.4	16.0 *	1.3

large long-term effect sizes). The probability of receiving a reward was positively related to (in particular unadjusted) long-term effect sizes (Pearson $r = 0.36$; $p < 0.001$). The length of the follow-up period was not related to any of the long-term effect sizes (Pearson $r = -0.004$; $p = 0.97$).

Furthermore, combinations of variables were entered in a number of GLM-analyses in order to find best fitting models for short- and long-term effect measures. The basic procedure in finding this model was to add one variable (or interaction) to the previous best fitting model (with the highest F-value). The model having the least number of degrees of freedom was preferred. Individual contributions of the variables to the model had to be statistically significant (SS2 and SS3). A summary of the results is presented in Table 5.4a–b.

Short-term effects Regardless of the type of dependent measure, the following variables consistently had a significant contribution: 'type of population', 'immediate and/or delayed delivery of rewards', and 'presence of a mandatory safety belt use law' (or 'baseline level of belt use'); these latter two variables were no longer significant when adjusted effect sizes were used as dependent variables. Adding more variables (e.g., 'group and/or individual rewards'; 'theory referred to') to the obtained models resulted in unstable models in the sense that a combination of other variables sometimes gave an almost equal fit. In general, models with weighted effect sizes resulted in a larger proportion of explained variance than with unweighted effect sizes, and models with adjusted effect sizes had a larger proportion of explained variance than unadjusted effect sizes (see Table 5.4a–b).

Long-term effects In general, other variables were included in the long-term effect models, and the 'best fitting' models accounted for much less variance than those for short-term effects. Different, 'interchangeable' models were found for both adjusted and unadjusted effect sizes: a combination of other variables sometimes gave an almost equal fit. The presence of a promotional item, and the absence of a control group were the most important variables for large unadjusted effect sizes; low usage rate during baseline and noncontingent rewards led to large adjusted effect sizes (see Table 5.4a–b). Note that some variables, such as 'type of population', which were significant when entered as a single factor in the GLM-analyses (see Table 5.2), were no longer significant in the long-term effect models with combinations of variables.

In this section the results of the analyses of the combined dataset were presented. Since the main factor 'coder' did not account for a meaningful part of the variance ($F(1,135) = 0.48$; $r^2 = 0.004$), and the results of the data of coder 1 and 2 separ-

ately were similar to the ones obtained with the combined data, we do not describe the results of these separate analyses.

Apart from using the weights 1–4 in the analyses, also other values have been used that satisfied the ordinal relationship. However, because other weight distributions did not lead to different results from those already described, these are not reported separately.

Table 5.4a: Best fitting models for short-term effects; degrees of freedom ($df_{1,2}$), F-values, significance level (p), and proportion of variance accounted for (r^2) are shown for weighted, unweighted and adjusted effect size measures.

(a) Short-term effect

		Weighted				Unweighted		
Variables		$df_{1,2}$	F	p	r^2	F	p	r^2
(a1)	Mandatory safety belt use law + type of population	5,135	18.60	***	0.419	16.63	***	0.392
(a2)	Mandatory safety belt use law + type of population + immediate and/or delayed reward	7,135	18.79	***	0.509	16.45	***	0.476
(a3)	Mandatory safety belt use law + type of population + immediate and/or delayed reward + type of pop. × Immed./delayed rew	14,135	13.63	***	0.614	10.98	***	0.562
(a4)	Mandatory safety belt use law + Type of population + Immediate and/or delayed reward + Type of pop. × Immed./delayed rew + Group and/or individual reward	16,135	13.43	***	0.645	10.67	***	0.591
Or:								
(a5)	Usage rate during baseline + Type of population + Immediate and/or delayed reward + Type of pop. × Immed./delayed rew + Group and/or individual reward	16,135	13.28	***	0.641	11.07	***	0.600
		Adjusted, weighted				Adjusted, unweighted		
Variables		$df_{1,2}$	F	p	r^2	F	p	r^2
(a6)	Type of population + immediate and/or delayed reward	6,135	19.68	***	0.478	18.51	***	0.463
(a7)	Type of population + immediate and/or delayed reward + type of pop. × immed./delayed rew	13,135	12.52	***	0.572	10.29	***	0.523
(a8)	Type of population + immediate and/or delayed reward + type of pop. × immed./delayed rew + theory referred to	16,135	14.21	***	0.656	13.27	***	0.641

Table 5.4b: Best fitting models for long-term effects; degrees of freedom ($df_{1,2}$), F-values, significance level (p), and proportion of variance accounted for (r^2) are shown for weighted, unweighted and adjusted effect size measures.

(b) Long-term effect

		Weighted				Unweighted		
Variables		$df_{1,2}$	F	p	r^2	F	p	r^2
(b1)	Promotional item + control group	2,115	18.49	***	0.248	24.03	***	0.300
(b2)	Promotional item + control group + contingency of rewards	3,115	14.45	***	0.281	17.85	***	0.325
(b3)	Promotional item + control group + contingency of rewards + probability of receiving reward	4,115	12.74	***	0.317	14.85	***	0.351
		Adjusted, weighted				Adjusted, unweighted		
Variables		$df_{1,2}$	F	p	r^2	F	p	r^2
(b4)	Usage rate during baseline + contingency of rewards	2,115	30.40	***	0.350	33.21	***	0.380
(b5)	Usage rate during baseline + contingency of rewards + probability of receiving reward	3,115	25.51	***	0.406	27.75	***	0.426
(b6)	Usage rate during baseline + contingency of rewards + probability of receiving reward + control group	4,115	23.69	***	0.461	24.09	***	0.465
(b7)	Usage rate during baseline + contingency of rewards + probability of receiving reward + control group + expected value	5,115	24.08	***	0.523	23.72	***	0.529

5.3.9 Bootstrap analyses

The GLM-models a4, a5, a8, b3 and b7 for weighted and unweighted effect size measures (as indicated in Table 5.4a–b) were entered in a bootstrap-analysis to investigate their stability. GLM-analyses were performed on 5,000 random samples (with replacement) of the entries in the combined dataset. The results of the analyses (see Table 5.5) show how often each variable had a significant contribution to the specified model; when a variable appeared significant ($p < 0.05$) in a vast majority of cases ($> 90\%$), the moderator was concluded to be stable.

The results confirmed earlier indications that for unadjusted short-term effect sizes (models a4 and a5) the variables 'type of population', 'immediate and/or delayed delivery of rewards', and the interaction of these two variables were quite stable in the short-term-effect-models, whereas the variable 'group and/or individual rewards' and 'baseline belt use' (or 'mandatory belt use law') were less stable. When the variable 'group and/or individual rewards' was removed from the models (cf.

model a3), the variable 'baseline belt use' or 'mandatory belt use law' was also significant in more than 90% of cases. The results of the bootstrap analysis for adjusted short-term effect sizes (model a8) showed that all variables appeared to be stable (see Table 5.5).

For the long-term-effect-models with unadjusted effect size measures (model b3), all variables appeared unstable except for 'control group'; with adjusted effect size measures (model b7), the variables 'baseline belt use', 'contingency of rewards', 'probability of receiving a reward', and 'control group' appeared stable, whereas the variable 'expected value' was less stable (see Table 5.5).

In general, for both long- and short-term effects, models using adjusted effect size measures appeared more stable than those using unadjusted effect size measures; and models using weighted effect size measures were more stable than those using unweighted effect size measures.

5.4 Discussion

This meta-analysis confirms the conclusion of earlier narrative reviews (e.g., Geller, 1984b; Geller et al., 1987; Hagenzieker, 1988, 1992a) that incentive campaigns to stimulate safety belt use generally lead to substantial short-term effects; and that long-term effects are smaller than short-term effects, but belt use during follow-up measurements after withdrawal of the incentive campaigns is generally higher than initial baselines. Whereas earlier narrative reviews did not quantify the effect sizes of incentive campaigns, the results of this meta-analysis indicate that the mean effect sizes of incentive treatments that were reported in the literature amount to 12.0 (unweighted mean 20.6) and 9.6 (unweighted mean 13.7) percentage points over baseline for short- and long-term effects, respectively.

When the results of the present meta-analysis are compared with those of Johnston et al. (1994), it appears that the short-term effect sizes as reported by them are comparable to the ones found in this research. Johnston et al. used unweighted effect sizes, and found a median of 17.6 percentage points, while the median unweighted short-term effect size we obtained was 19.5 percentage points (mean 20.6). Most studies included in Johnston et al.'s analysis are also used in this analysis. In addition, we selected a number of studies that were not used by them (e.g., Foss, 1989; Hagenzieker, 1989, 1991; Hunter et al., 1991a; Wojtowicz et al., 1992). Moreover, we included studies involving children if the safety belt usage of (their) parents was measured, whereas Johnston et al. excluded those. In all, Johnston et al. included 20 articles with 36 effect measures in their analysis, compared to 34 articles with – on average per coder – 69 effect-sizes in our study. Johnston et al. did not

Table 5.5: Summary of results of the bootstrap analyses; for each variable the percentage of cases is shown (% signif.) – out of 5,000 bootstrap analyses – that resulted in a significant contribution to the various GLM-models (a4, a5, a8, b3, and b7; see also Table 5.4a–b).

Short-term effect	Unadjusted (model a4)		Unadjusted (model a5)		Adjusted (model a8)	
	Weighted % significant	Unweighted % significant	Weighted % significant	Unweighted % significant	Weighted % significant	Unweighted % significant
Mandatory safety belt use law	82	55	–*	–	–	–
Usage rate during baseline	–	–	82	68	–	–
Type of population	100	100	100	100	100	100
Immediate and/or delayed reward	97	94	97	94	99	98
Type of pop. × Immed./delayed rew.	99	97	100	98	100	98
Group and/or individual reward	82	73	78	67	–	–
Theory referred to	–	–	–	–	97	99

Long-term effect	Unadjusted (model b3)		Adjusted (model b7)	
	Weighted % significant	Unweighted % significant	Weighted % significant	Unweighted % significant
Usage rate during baseline	–	–	100	100
Promotional item	47	82	–	–
Control group	95	92	98	91
Contingency of rewards	74	71	95	96
Probability of receiving reward	87	78	99	99
Expected value	–	–	85	82

* '–' : variable not in specified model

differentiate long-term effect sizes by type of intervention (e.g., enforcement, incentive, education). Nevertheless, they found an overall long-term effect size of 13.0 percentage points, which is very similar to the unweighted long-term effect size we obtained (an increase of 13.7 percentage points compared to baseline). So, it appears that even though the number and type of studies included in these two meta-analyses differ, as well as the exact method by which the meta-analyses were conducted, the overall results are very much alike and therefore strengthen the conclusion that considerable effects of incentive campaigns on safety belt usage can be reached.

At the same time, our findings reveal that the magnitude of the (short-term) effects depend on a number of moderator variables: the effect is not reached 'no matter what' type of incentive campaign is implemented. The results of this meta-analysis show that the short-term effect size is moderated by a number of single variables, and that a combination of the following variables accounted for 64% of its variance:

(1) The type of population involved; campaigns at elementary schools showed the largest effect sizes, which finding is in line with previous reviews that conclude that incentive programs implemented in small, homogeneous groups give better results than those in large, heterogeneous groups such as communities.

(2) The immediacy of delivering the rewards; when incentives were delivered immediately, larger effect sizes were reported than when incentives were delivered delayed; this finding was not reported in earlier reviews on the effects of incentive programs to stimulate safety belt use, but is in line with (learning) theory (see, for example, Bandura, 1986).

(3) The initial baseline rate (which was highly correlated with the presence or absence of a safety belt usage law); lower initial baselines and absence of a safety belt use law yielded larger effect sizes (see also, Elliott, 1993).

When used as the only explanatory variable, these variables could explain a substantial part of the variance of effect measures, regardless of the type of effect size measure (e.g., weighted, unweighted). However, as an exception, the initial baseline level could not explain a substantial part of the variance of any of the adjusted effect measures. This suggests that it is indeed 'more difficult' to increase belt use by a given number of percentage points starting at higher baselines.

Incentives that were based on group behaviour yielded larger short-term effect sizes than when they were based on individual behaviour. However, the results of the bootstrap analyses showed that this variable was less stable than the other variables in the models for short-term effect sizes. This may indicate that the relevance

of this variable was dependent on a small number of studies. The variable 'theory referred to' was coded in order to explore whether programs that were 'theory driven' would be more effective than those that were not (see, for example, Nelson & Moffit, 1988); it was found that, when reference was made to motivational theories, reported effect sizes were larger than when no theory, or another theory, was referred to. Also, when promotional items or immediate valuables and noncontingent rewards were present larger effect sizes were reported. And when a control group was included in the research design, smaller effect sizes were on average reported than when no control group was present (cf. Elliott, 1993). However, when these variables were included in a model together with other variables, their impact to differentiate between effect sizes disappeared. This suggests that variations of these characteristics coincide with variations in other characteristics (such as the type of population involved, the size of the study etc.), making it difficult to isolate separate contributions to the explanation of variance.

Other variables, such as the duration of the intervention (see also Johnston et al., 1994), the probability of receiving a reward, and the value of the reward – and also the expected value of the reward – appeared unrelated to the short-term effect sizes. However, the latter three variables were difficult to code, and could therefore often be coded incorrectly, which might be a reason for the absence of a statistical significant relationship with effect size.

The results for mediating variables explaining the variance in the long-term effect sizes were less clear. Although a number of variables explained a significant part when entered as single variables in a model, they appeared insignificant in models with combinations of these variables. The presence of a promotional item, and the absence of a control group in the research design were the most important variables for unadjusted large long-term effect sizes. The results of the bootstrap analyses indicate that the variable 'control group' appeared to be the most consistent stable variable across various long-term effect size models. Furthermore, low usage rate during baseline and noncontingent rewards seem to result in large adjusted effect sizes. At first sight, this latter finding is rather puzzling because it was expected that contingent rewards would lead to larger effect sizes than noncontingent rewards (see, for example, Johnson & Geller, 1984; Bandura, 1986). However, this finding could be an artefact of the very unstable long-term-effect models.The confounding of small scale studies, the absence of promotional items or immediate valuables, noncontingent rewards, and large long-term effect sizes has possibly resulted in coincidental significant contributions in some of the hypothesized models. The length of the follow-up period (since ending of program), and the value of the reward had no effect on the long-term effect size. The probability of receiving a reward was positively related to long-term effect size, but was rarely significant when entered with other variables in a model. Also the type of reward, other than promotional items, did not influence the long-term effect. We do not have

a clear explanation for the discrepant results for long-term effects as compared to those for short-term effects. It is probable that the way long-term effects have been measured differs so much between studies that this hampers further analysis.

Based on the results of the best fitting models, the largest effect sizes of incentive campaigns to promote safety belt use can be obtained when the population involved consists of elementary school pupils and their parents, when incentives are delivered immediately rather than delayed, and with relatively low baseline levels (and an absence of a mandatory safety belt use law). On the other hand, the smallest effect sizes are obtained for incentive campaigns in community settings with relatively high baseline levels; it seems that for this type of conditions another approach, such as selective enforcement in combination with publicity might be more efficient in enhancing safety belt use (see, for example, Zaal, 1994 for a review of traffic law enforcement research findings).

Finally, a number of limitations of the present study should be mentioned. First, although we have attempted to include all relevant studies – and exclude irrelevant ones – in this meta-analysis, it is still possible that we missed certain studies or incorporated studies that should not have been included according to our criteria. For example, the results of evaluation studies were often reported in more than one publication or (parts of) results (of follow-up measurements) of campaigns were distributed over various publications; it was sometimes difficult to assess whether particular findings were already present in our database. Connected with this is the issue of the independency of studies: although the analysis assumes independence, this was actually not always the case. For example, the coding of the variable 'follow-up study?' shows that about one fifth of the studies in the dataset was preceded by another study (most often also included in this meta-analysis) of an incentive program in the same population. However, when entered in the analyses, this variable did not significantly contribute to the explanation of the variation in effect sizes. It appeared that more than one intervention in the same population did not lead to discrepant effect sizes. Furthermore, the aim of the meta-analysis was to determine the effect of campaigns that only consisted of incentive programs. Therefore, incentive campaigns that also had an enforcement component were excluded from the analysis. However, the contents of the campaigns were sometimes vaguely described, which might have led to erroneous in- or exclusion. For instance, the well-known incentive campaign as described by Spoonhour (1981) appeared to have had an enforcement component as well, which only became apparent from another publication (Geller, 1982).

Second, as Miller & Pollock (1995) emphasize, the outcomes of meta-analytic studies are particularly fragile when the effects of moderator variables remain unexamined. In the present meta-analysis a large number of moderator variables were included, both methodological and substantive variables, the latter primar-

ily based on findings reported in earlier (narrative) reviews (e.g., Geller, 1984b; Geller et al., 1987) and on theoretical notions (see, for example, Bandura, 1977, 1986; Festinger, 1957; Lepper & Greene, 1978). However, other moderator variables may exist; obviously, the possible effects of mediating variables that were not coded are unknown. For instance, only tangible reward characteristics of the campaigns were coded; it might be possible that other campaign characteristics, such as the type and amount of publicity material, or certain educational aspects can explain the effect size variation better.

Thirdly, zero or negative effects may not be reported in the literature. Indeed, it appeared that large short-term effect sizes were somewhat over represented, and that only few large studies (which have relatively small effect sizes) were present in the dataset. On the other hand, studies in this meta-analysis that were published in *journal articles* did on average not report larger effect sizes than those reported in *research reports*.

Furthermore, although incentive campaigns have been extensively researched, many studies did not report sufficient (detailed) information for the calculation of effect sizes, and sometimes effect sizes had to be derived from graphs. For example, standard deviations – as commonly used in effect size measures – were almost always missing in the reviewed reports. Therefore, alternative effect size measures were chosen and were weighted with (an estimate) of the number of observations on which they were based. By using various effect-size measures and weight distributions, as wel as by conducting homogeneity and bootstrap analyses in our meta-analysis – which is, as far as we know, seldom if never done before in meta-analysis – we have attempted to take into account the unreliability of the data and assess the sensitivity and robustness of the results. It appears that this approach has been fruitful, since it helps to gain insight in the possible range of effect-sizes that can be obtained instead of focussing on one single effect-size estimate, and also clearly reveals confoundings of variables.

In addition, the descriptions of the interventions themselves were often minimal and led to discrepancies in coding because the required information had to be inferred from 'circumstantial evidence'. Some variables appeared very difficult to code, because information was not present in the articles and estimates or 'expert guesses' had to be made, for example, concerning the probability of receiving a reward, the value of the reward, the observation method used (obtrusive or nonobtrusive), etc. For this reason, the data of the two coders were both analysed in combination and separately, to investigate whether differences in coding would lead to differences in outcomes. Despite discrepancies between coders, a.o. in the number of treatments identified, correlations between coders were generally quite high and when entered in the analyses no *systematic* variation between coders was found.

Finally, the impact of research methodology should be mentioned, such as the inclusion of a control group in the research design, the number of observations on which the reported effect sizes were based, and the observation method employed. Usually effect sizes are corrected for controls. However, only a minority of studies in the present meta-analysis had included a control group. Possibly, the applied nature of the studies which makes it often difficult to find control sites, and/or financial reasons associated with the evaluation of interventions in applied settings, have contributed to the absence of control groups in the research design. This is a problematic aspect, because studies without controls appear to report on average larger effect sizes than those without (e.g., Lipsey & Wilson, 1993; Elliott, 1993). Lipsey & Wilson (1993) report 61% larger effect sizes when no control group is present, a finding which is comparable to our findings. However, the presence or absence of a control group did not statistically contribute anymore to the explanation of variation in the effect sizes when this variable was used in combination with other moderator variables. Because larger studies (with many observations) more often involve control groups and on average report smaller effect sizes, it is virtually impossible to disentangle the effects of these two factors. In general, it appears that the primary research on the effects of incentive campaigns to promote safety belt use contains some blind spots caused by confounding variables. Many characteristics are not sufficiently varied across studies to be able to estimate their individual contribution to the overall effect size.

In summary, the technical quality of the reported studies that have evaluated the effects of incentive campaigns is not as good as one would like it to be. However, despite the limitations of the present meta-analysis, the results appear to be quite robust as far as the short-term effects are concerned. We feel that the results of the meta-analysis have added to our knowledge of the impact of incentive campaigns to stimulate safety belt use, both in quantifying the (range of) effect sizes that can be obtained and in locating (combinations of) variables that mediate the effect sizes; findings that have not become apparent in previous (narrative) reviews on the same topic.

Chapter 6

Summary, discussion and conclusions

It has been traditional practice in the traffic system for desired behaviour to be stimulated by rules, regulations, and laws, resulting in the punishment of offences. Until recently, modifying road user behaviour by reward has not been given much attention in the practice of road safety. In this monograph it was aimed to get a better understanding of the possible role of reward in influencing road user behaviour. To this end, some theoretical notions related to the assumed effects of rewards on behaviour were considered, as well as the role of rewards in other fields of application, and an overview of past studies that have been conducted in this area was presented. The main questions that were addressed are whether rewards are effective in modifying individual road user behaviour, which types of reward (programs) are more effective than others, which other conditions or circumstances mediate this effectiveness, and whether reward programs can be successfully applied on a larger scale.

Firstly, this concluding chapter recapitulates the theoretical contributions and empirical findings of reward activities to stimulate safe road user behaviour. The effectiveness of reward programs in the field of road safety is assessed based on a theoretical analysis and empirical research, including a meta-analysis (Section 6.1). Next, in Section 6.2, it is discussed what the large-scale applicability might be for reward programs on road user behaviour. In Section 6.3 a general framework is attempted for this type of research and guidelines are formulated tentatively for setting up and implementing effective reward programs. In the final Section, 6.4, the issue of theory- and practice-driven research on road user behaviour is addressed.

6.1 Summary: Effects of rewards on road user behaviour

Are reward programs effective in changing (road user) behaviour, and if so, under what circumstances and to what extent?

In *Chapter 1* the results of a literature review are reported in which attention is paid to the role of rewards in psychological theories and models of road user behaviour. In addition, an inventory is made of what is known – as becomes apparent from evaluation studies – about the effects of rewards on road user behaviour. A multitude of psychological theories – generally in the field of learning and motivation – reserve an important place for the principle of reward. In theories varying from Skinner's operant conditioning and Bandura's social learning theory to theories with a more cognitive orientation, such as Festinger's cognitive dissonance theory and (subjective) utility theories, rewards are explicitly regarded as powerful influences on behaviour, even though the principles and presumed mechanisms of effect of the individual theories can vary greatly. While in learning theories rewards are generally ascribed a main role in changing behaviour, social cognitive oriented theories emphasize the interaction of rewards with cognitive processes such as (intrinsic) motivation. From the various approaches, variables such as the size and type of the reward, the probability and timing of rewarding, the type of task or behaviour one is rewarded for, the social setting, cost-benefit considerations and other cognitive evaluations of these variables can be identified as important factors mediating the effects of rewards on behaviour.

In models of road user behaviour the concept of reward has been given an explicit role much less often. In two of these models (Wilde's risk homeostasis theory and Janssen's utility function model), effective rewards are assumed to be contingent on a certain outcome (no accidents); in another model (Fuller's threat avoidance model) rewards are assumed to be helpful in adjusting certain unsafe behaviours that are experienced as pleasurable or exciting (such as speeding).

In various fields of application rewards have been successfully used to stimulate desired behaviour. For example, to motivate employees in industry to adhere to safety regulations, to promote environment-relevant behaviours, and to stimulate certain preventive health-precautionary behaviours. Such reward programs have usually been applied in company settings. These settings seem to be a good environment for the application of reward programs, because these can be cost-effective for the company. If workers demonstrate safer behaviour on the road, for example, this could reduce absenteism and (indirectly) increase productivity, as fewer accidents occur. These applications usually concern more or less homogeneous populations and a small-scale setting, which facilitates the practical implementation. Examples are also available of somewhat larger-scale reward programs for

accident-free driving. The research available on the effects of rewards for specific safe behaviour in traffic has concentrated on the efforts to stimulate the use of safety belts. Such reward programs general lead to substantial increases in safety belt use. Problematic is that most of these reward programs have been applied in conditions of voluntary safety belt use and (hence) with relatively low baseline levels of safety belt use, and that reward program characteristics have not been systematically varied. This makes it very difficult to determine which (combinations of) factors are optimally effective, and whether reward programs can also be effective in conditions with relatively high baseline levels of safety belt use.

In order to get more insight in the obtainable effect sizes and factors mediating the effect of reward programs, one specific discrete behaviour was selected – safety belt use – to concentrate further studies on. A number of field studies have been conducted in which the effect of various reward programs on the use of safety belts was investigated in situations with relatively high baseline levels of belt use (Chapters 2-4). In addition, the effects of 'rewards' as compared to police enforcement were investigated. Furthermore, it was examined – by means of a meta-analysis – how the results of these reward programs relate to previously reported results, and which characteristics of reward programs can be determined that mediate the effectiveness of such programs (Chapter 5).

The small-scale study reported in *Chapter 2* was set up to explore whether a reward campaign aimed at increasing safety belt use could also be effective under conditions of mandatory safety belt use. The results of this study showed no significant increases in observed safety belt use directly after the campaign. Contrary to the general results reported in the literature, not even a short-term increase was observed. Possible explanations for the absence of an effect that were offered were that the reward program was not sufficiently known by the public, that rewarding only once and on a single day – as was the case here – was not enough to induce behaviour change, and that the reward offered was not experienced as an 'attractive' one.

As a follow-up, a larger-scale field study was conducted at twelve different military bases in the Netherlands *(Chapters 3 and 4)*. Amount of enforcement, type of publicity, and reward strategies were varied between military bases. To overcome some of the short-comings of the pilot study, considerable effort was put in publicizing the campaign, it was sought to offer attractive rewards for the target group (for example, prizes included money coupons and compact disc players), and the campaign had a longer duration (two months) during which period several prize drawings took place. In addition, different reward strategies (individual vs. group-based rewards, variation in number of rewards) were compared. Moreover, the effects of the reward strategies implemented at some military bases were compared to the effects of police enforcement at other bases.

It turned out that both enforcement and reward programs had the same overall effect. Enforcement and reward programs showed mean increases from 67% and 62% during baseline to 75% and 69% directly after the campaign, and to 78% and 76% during follow-up, respectively. So, even when behaviour is regulated by law rewards are effective in increasing the (already mandatory) behaviour. An important finding was that reward programs were capable of enhancing safety belt use beyond high initial baseline belt use rates. This holds especially for the individual reward programs. The frequency of rewarding did not lead to differential effects: offering one prize per week was equally effective as four prizes per week. The group-dependent reward program – involving competition between two bases – was less effective.

In addition to behavioural observations, the campaign was also evaluated by means of a written survey administered to the personnel of the military bases. The survey contained items concerning reported belt use, motivation to use a safety belt, attitudes toward legislation, public information, enforcement and reward strategies, and awareness of the campaign. Contrary to results obtained elsewhere, this study showed that rewards were not rated among the most popular countermeasures. The results showed that enforcement was clearly a better accepted countermeasure than rewarding drivers for the use of safety belts. The fact that personnel of military bases were the target group could be a reason for this finding. Respondents from bases exposed to rewards tended to have relatively more positive opinions than those exposed to enforcement. However, the not-so-positive opinions of rewards did not result in a smaller effect on actual safety belt use than the more accepted enforcement treatment. The notion that tangible rewards might have a negative effect on intrinsic motivation does not seem to be supported by the findings of this survey: The most important motives to use a safety belt as indicated by the respondents were: because of its protective properties in case of crashes, habituation, and because it is compulsory. Only a small minority of respondents indicated that they use their safety belt (more often) because of a reward (although respondents from the incentive bases reported about twice as often that rewards are an important motive to use a safety belt as compared to those from enforcement bases).

To date, quite a number of empirical studies have been performed to assess the short- and long-term effects of reward programs in stimulating safety belt use: what may be concluded from these studies, and is it possible to identify moderating variables that are related to the magnitude of these effects? The review based on a meta-analysis (*Chapter 5*) was confined to studies that had used behavioural observations of safety belt use, and to programs that included rewards that could be expressed in terms of some ‘material value’.

The results of the meta-analysis confirmed the conclusion of earlier narrative reviews that reward campaigns to stimulate safety belt use generally lead to substantial short-term effects; and that long-term effects are smaller than short-term effects, but belt use during follow-up measurements after withdrawal of the reward campaigns is generally higher than initial baselines. Whereas earlier narrative reviews did not quantify the effect sizes of reward campaigns, the results of this meta-analysis indicate that the mean effect sizes of reward programs that were reported in the literature amount to 12.0 (unweighted mean 20.6) and 9.6 (unweighted mean 13.7) percentage points over baseline for short- and long-term effects, respectively.

The magnitude of the (short-term) effects turned out to depend on a number of intervening variables of which the following three variables accounted for much of its variance:

- The target group involved; campaigns at elementary schools showed the largest effect sizes, followed by campaigns at industrial/company settings. This finding is in line with previous reviews that conclude that reward programs implemented in small, homogeneous groups give better results than those in large, heterogeneous groups.
- The time of delivering the rewards; when rewards were delivered immediately, larger effect sizes were reported than when rewards were delivered delayed; a combination of both immediate and delayed rewards showed the largest effect sizes. This finding was not reported in earlier reviews on the effects of reward programs to stimulate safety belt use, but is in line with (learning) theory.
- The initial baseline rate (which was highly correlated with the presence or absence of a safety belt usage law); lower initial baselines and absence of a safety belt use law yielded larger effect sizes.

It was also found that reward based on group behaviour on average yielded larger short-term effect sizes than when they were based on individual behaviour; a combination of both individual and group rewards showed the largest effect sizes. However, mixed results of this variable have been reported, and the effect of this variable was less stable as compared to other variables. It should also be noted that variations of these characteristics generally coincide with variations in other characteristics (such as the type of population involved, the size of the study etc.), making it difficult to isolate separate contributions to the explanation of variance. It also turned out that even campaigns of short duration, with low probabilities of receiving a reward and with prizes of little monetary value were effective; these

variables appeared unrelated to the short-term effect sizes. Finally, the results for mediating variables explaining the variance in the long-term effect sizes were less clear. It was concluded that the way long-term effects have been measured probably differed so much between studies that this hampered further analysis.

General conclusions

Based on the results of these and previous studies as described in this monograph the following general conclusions with regard to the effectiveness of rewards to stimulate safety belt use can be formulated:

- Rewards are more effective with relatively low baseline levels and in the absence of a mandatory safety belt use as compared to conditions with relatively high baseline levels and mandatory safety belt use.
- Rewards can be effective under conditions of mandatory safety belt use with relatively high baseline levels of belt use. Furthermore, this effect is comparable to that of police enforcement.
- Reward programs are most effective in small scale settings; larger-scale applications are less effective but still show significant increases in safety belt use.
- Reward programs are especially effective in reaching substantial short-term effects.
- Long-term effects are smaller than short-term effects, but belt use during follow-up measurements after withdrawal of the reward campaigns is generally higher than initial baselines.
- Immediate delivery of rewards is generally more effective than delayed rewards; a combination of both appears to be most effective.
- Both individual and group oriented programs can be effective. On average, group based rewards show larger effect sizes than individual rewards, and a combination of both seems to be most effective.
- The size (value) of the reward, the probability of the reward, and the duration of the reward campaign are not related to the magnitude of the effect.

6.2 Large-scale applicability of reward programs in road traffic

Reward programs have shown to be effective and applicable in the area of road safety, as became evident from evaluation studies. Can reward programs, however, also be applied in a more structural, large-scale manner?

It has been argued earlier that a number of advantages of reward programs can be named in comparison to other approaches not involving rewards. For example, whereas it has been shown that information and media campaigns often do not lead to behaviour change, reward programs have shown to be effective in accomplishing substantial behaviour change, in particular in the short term. Furthermore, reward programs may be less costly than enforcement programs, requiring comparatively less manpower and administrative actions when applied in small-scale environments. In industry, reward programs generally have appeared to be cost-effective. Another advantage of rewards as compared to 'punishment' is that reward programs are often considered as measures that are generally rated favourably among the public, as questionnaire studies have shown, notwithstanding that at least one survey showed that (a group of Dutch people) clearly preferred 'punishment' (enforcement) for the non-use of safety belts to the promise of rewards for using a safety belt (see Chapter 4). Finally, rewards have the advantage over punishment for those (dangerous) behaviours that are not explicitly described by law or that are difficult to check, and are therefore difficult to enforce. Dangerous overtaking and too close following, for example, are behaviours that are difficult for the police to enforce because the proof is difficult to provide. Because rewards are not part of the complicated judicial system, rewards can possibly help to change such behaviours.

However, some problem factors can be anticipated with regard to how rewards can be structurally integrated in the traditional traffic system. These concern the type of behaviour to be rewarded, the nature and scale of the reward program, the necessity to systematically and carefully monitor the behaviour concerned, and the organisation and agents that set up the program and deliver the rewards, respectively.

Because rewards have proven to be effective, it can be argued that it would be 'unethical' not to use them, although some have different views about this. The issue of rewarding road users for desired behaviour has been criticized by those who reason that such behaviour should be considered as part of 'normal' road user behaviour, and should therefore not be rewarded; instead deviations from the normative behaviour should be 'punished'. The objection seems more to how the behaviour is being influenced than to the fact that it can be influenced by rewards. In addition, 'punishment' (police enforcement) is built-in into the traffic

system, whereas rewards are not. Reward programs require careful monitoring of the behaviour concerned. Large-scale applications require that reward programs be implemented in the traffic system, and some organisation or agents should be responsible for the monitoring of behaviour and the delivery of rewards. However, one can be hesitant about government bodies actually rewarding such desired behaviour. These are in fact the same bodies that are supposed to punish breaking the law. One possibility is that private companies or organizations are used to encourage desirable road user behaviour by offering rewards; these could also be the agents that deliver the rewards. For example, insurance companies could offer incentives for specific (safe) behaviours, such as using a safety belt. However, insurance companies do not appear to be willing to introduce such incentives. Other reward systems that are currently applied in the form of premium discounts that car insurers offer their clients have some drawbacks. For example, these are often not individual rewards in the sense that the discount is determined on a statistical basis for certain population groups. And when discounts depend on the number of years of accident-free driving, these discounts are usually not progressive with the year-to-year accumulation of an accident-free record, and after a certain period of time a 'forgiveness clause' comes in effect which saves drivers from an increase in their insurance fees if they have an accident. Furthermore, it is possible that the loss of a premium discount as a result of submitting a claim is more likely to be regarded as a punishment than its retention is regarded as a reward by the insurance holder. Seen from this perspective, a more effective form of reward would be to repay part of the premium for accident-free driving in a certain period; this is indeed found to be effective in the one study available on this subject (Vaaje, 1991).

In conclusion, structural large-scale implementation of reward programs is not easy but nonetheless worthwhile to study further, because of the substantial effects found in many studies. Although reward programs have been most successful in small-scale settings, both in the area of road safety as well as in other fields of application, some larger-scale applications have also shown significant effects.

Now, after reviewing the research reported and addressing the issue of large-scale applicability, one may question whether there is a lesson to learn from the results for setting up a general framework to study the effectiveness of reward programs and to formulate guidelines for such programs, be it only tentatively. An attempt is made in the next Section.

6.3 Framework and tentative guidelines for effective reward programs

As a general framework for studying the effects of social interventions - and more specifically also, reward programs – may serve the basic evaluation model defined as a model in which persons (or other observation units) that the reward program is expected to affect have been assigned to levels of the intervention, and subsequently some outcome variable or variables are measured that are expected to show the impact of the intervention (adapted from Judd & Kenny, 1981, p. 9; Shadish et al., 1991, discuss program evaluation and the different approaches proposed more thoroughly). This approach does not only comprise the (methodological) set up of evaluation studies per sé, but also incorporates the contents of the specific programs evaluated. What are the crucial elements of this basic evaluation model, what is the framework behind it and can guidelines be formed to fill in this framework?

The crucial elements for the basic evaluation model – the foundations of the framework – are: the units of observation (e.g., persons or subjects), the intervention (viz. reward programs) and the outcome variables (viz. safety belt use or accident-free driving). Sure enough, this basic model for program evaluation can easily be augmented by explicitly taking into account different settings and time of estimating the effects of the intervention, or by specifying in more detail each of the model's constituent elements. For our present purpose it suffices to stick to these distinctive features.

Guidelines for effective programs have to do with each (or a combination) of the above constituent parts of the program evaluation model. Which guidelines can be formulated from earlier research findings, or are possibly implied by theoretical approaches (as described in Chapter 1)?

Based on the available knowledge derived from several theoretical approaches and applications in the area of road safety as well as in other fields, which have been described in this monograph, the following tentative guidelines for effective reward programs can be formulated. The guidelines are embedded in the general framework of the basic evaluation model and may be used as an aid for applied researchers or practitioners, and are an elaboration of earlier proposals (see, for example, Campbell, 1991; Geller, 1982, 1992; Hale, 1991; Lonero et al., 1994; Streff & Geller, 1986; Wilde, 1994; Winett et al., 1989).

The guidelines apply to both behaviour- and outcome-oriented reward programs. In the latter type of reward program the outcome and not the mediating behaviours leading to that outcome are rewarded. However, this does not mean that these

mediating behaviours are irrelevant or should remain unspecified. It is recommended that also in outcome oriented reward programs the participants involved be informed about how – by which behaviours – they could achieve this outcome. Except for Wilde (1994), who specifically recommends only to reward outcomes – not behaviours –, others who have made up lists of guidelines for reward programs recognize that also in outcome-oriented reward programs it is important to pay attention to the behaviours leading to that outcome (see, e.g., Campbell, 1991; Geller, 1982, 1992; Hale, 1991; Streff & Geller, 1986; Winett et al., 1989).

As has become apparent in the previous chapters of this monograph, there are still many questions as how to design the 'ideal' reward program because in evaluations of reward programs the characteristics of these programs have not been systematically varied. The list presented here should be seen as an attempt to provide concrete guidelines which are sometimes based on empirical findings, and sometimes primarily on theoretical considerations. In Table 6.1 for each guideline it is assessed whether it is mainly based on theoretical considerations (indicated by 'o') or on empirical evidence from evaluation studies; '+' and '++' indicate if the guideline is backed up by 'some' or 'strong' empirical support, respectively.

Table 6.1: Tentative guidelines for effective reward programs, and assessments whether these are based on theoretical considerations (o), or on some (+) or strong (++) empirical support.

Guideline	Assessment
Type of behaviour (outcome variables)	
• the targeted behaviour should preferably be a discrete ("on/off") behaviour, such as using a safety belt	++
• the targeted behaviour should be specific ("use a safety belt") rather than general ("drive safely")	o/+
• reward programs for continuous behaviours (such as speed behaviour) are generally less effective	o/+
• the easier it is to reliably measure performance of the behaviour, the more effective the reward program	o
• if an outcome-oriented reward program is chosen, specify through which behaviours the outcome can be achieved	o
Participants and target group (units of observation)	
• the target group should have or be given the necessary knowledge and skill to exploit the contingencies	o
• one should know exactly how to behave and how this is monitored in order to get a reward, i.e. the reward-contingencies should be simple and clear	o
• targets and rewards should preferably be set in consultation with the participants	o
• the participants should believe that the attainment of rewards is dependent on their own behaviour	o

Table 6.1: Tentative guidelines for effective reward programs. (continued)

Guideline	Assessment
Reward program characteristics (intervention)	
• avoid unexpected, unanticipated rewards; these can have adverse effects	++
• rewards do not have to be delivered every time the target behaviour is displayed	++
• use direct and immediate rewards; these can be supplemented by a lottery system	++
• the programs should preferably include not only individual rewards but also group-based rewards; however, avoid competition between groups or individuals	++
• provide clear and repeated information and publicity about the reward campaign; the participants should not only be informed about the program in existence, but they should also frequently be reminded of it	+
• rewards should be perceived as attractive by the participants	+
• rewards do not have to be very large to be effective; the reward should be large enough to induce behaviour change, but not so large that it can be reasoned that it was the sole motivator	+/o
• a reward program can be more effective when used in combination with other interventions, such as training or police enforcement	o/+
• specific, detailed and realistic targets should be set	o
• rapid and clear feedback about progress to the goals enhance the effects	o
• rewards should be attainable and progressive; one failure should not lead to all subsequent rewards being unattainable, but continued success should be more highly rewarded as it nears the target	o
• information about the program should not only contain 'the rules of program' but should also make clear why – apart from the rewards – the target behaviour is personally relevant for the participants (safety relevance)	o
• make sure that participants see that others are being rewarded	o
• the reward program should be perceived as equitable by the participants; those who are not eligible for a reward should not resent this, and those rewarded should be seen by others as justly receiving it	o
Setting	
• select a relatively small, homogeneous target group; appropriate settings are, for example, companies, schools, and local communities	++
• carefully select the 'agents' that deliver the rewards; these should preferably be perceived as 'important' and 'respected' by the participants; appropriate agents could be, e.g., police officers (in community programs), teachers (in school-based programs), management (in company-based programs).	o
Time	
• reward programs do not have to have a long duration to be effective; programs with a duration of several weeks are effective in reaching substantial short-term effects	++
• systematic monitoring of the target behaviour is necessary	+
• measure the target behaviour also before the program is implemented in order to set realistic targets and to be able to determine its effectiveness afterwards	+
• in order to obtain long(er) term effects the program should be repeated regularly	o/+

6.4 Theory-driven or practice-driven research

After reviewing the research results on the effects of rewards on road user behaviour and considering theoretical issues in this respect, one may question whether there is a lesson to learn for developing behavioural models and for theories of road user behaviour.

The research on the effects of rewards on road user behaviour can be said to be mainly 'practice-driven'. Reward programs are often tried on an ad hoc basis as a solution for a specific problem. Although such attempts are inspired by theoretical notions, they are not theory-driven. Theoretical considerations are hardly explicitly taken as a starting point for investigating the effects of rewards on road user behaviour. Reference is sometimes made to particular theoretical approaches, such as learning theories or social cognitive theories, but specific hypotheses derived from these (classes of) theories have generally not been tested in the applied studies. The problem is that 'basic' psychological theories that pay attention to the role of rewards are difficult to translate to complex behaviours in an applied setting, such as road user behaviour. On the other hand, models of road user behaviour generally do not pay attention to the possible role of rewards. It is perhaps therefore that much of the work studying the effects of rewards on road user behaviour seems to be inspired not so much by models of road user behaviour, but by principles derived from basic psychological theories.

Although a number of variables that are considered theoretically important for the effectiveness of reward programs have also been studied in the area of road safety, the result of these studies is a collection of empirical findings which are difficult to generalize, because these are not embedded in an overall theory. For example, some variables that are considered to be important in a number of theoretical approaches, such as the size of the reward and the probability of rewarding, do not appear to moderate the effectiveness of reward programs to stimulate safety belt use. On the other hand, for example immediate rewards are more effective in increasing safety belt use than delayed rewards, as operant conditioning would predict; and group-based rewards are generally more effective than individual rewards, which is in line with social learning theory. There are also indications that group-based rewards that involve competition are not effective, which is compatible with social learning theory and approaches that stress the issue of intrinsic motivation. However, evaluations of reward programs did not always have adequate research designs, and results are often based on company reports. For example, control or comparison groups are often not included in the research design. Possibly, the applied nature of the studies which makes it difficult to find control or comparable sites, and/or financial reasons associated with the evaluation of interventions in applied settings, have contributed to the absence of control/comparison groups in the

research design. This is a problematic aspect, because studies without controls appear to report on average larger effect sizes than those without. For these reasons, firm conclusions usually cannot be drawn.

Three examples of unanswered questions further illustrate the difficulties of interpreting empirical findings that are not embedded in an overall theoretical framework and vice versa (theoretical framework that is not yet being tested properly).

1. The issue whether outcome- or behaviour-oriented rewards are more effective to achieve safer road user behaviour is far from solved. Empirical results so far indicate that both types of rewards are effective in the sense that positive effects have been reported on the chosen dependent variable – either some specific behaviour or in terms of outcomes resulting from unspecified behaviours. The relation between these types of reward program, however, has not been the topic of empirical study. Although strategies of rewarding safe outcomes appear to lead to improvements in road safety, we do not really know through what changes in behaviour such improvements come about. The identification of the mechanisms which mediate such improvements will almost certainly lead to the development of more effective and more efficient interventions. Thus far we have little more than a theoretical framework from which hypotheses may be derived about what is happening (Fuller, 1991b).

2. Which type of behaviour can or cannot not be effectively rewarded? This question remains unanswered. It has been argued earlier that rewards for modifying discrete road user behaviours (such as wearing a safety belt) can be assumed to be more effective than those for continuous behaviours (such as speed behaviour). The research available on the effects of rewards for specific safe behaviour in traffic, however, has concentrated in a somewhat one-sided manner on the efforts to stimulate the use of safety belts. As a result, conclusions concerning the effects of reward campaigns on other behaviours than safety belt use cannot be drawn.

3. Also the effect of rewarding in relation to other (motivational) measures, such as police enforcement, media campaigns or education, is unclear. We do not know which (combination of) measures is most effective and efficient for which types of behaviours and in which conditions. From the available results it can, for example, be concluded that besides rewarding desired, but voluntary behaviour, also rule following can be motivated by the introduction of rewards. However, the boundaries and limitations of rewards as a measure to stimulate desired road user behaviour are far from clear. A starting point could be a framework that integrates various (basic) theoretical

approaches with practical issues, in order to select appropriate measures for different types of behaviour (see, e.g., Vlek et al., 1997).

These difficulties not only apply to the research on rewards to influence road user behaviour. As Huguenin (1997) points out, the development of traffic psychology as a whole is characterized by the fact that in the past 25 years a large number of experiments and investigations were carried out to provide answers to individual questions and solutions for parts of a problem. This situation is probably not due to an indifference towards theories. It is certainly related to the complexity and heterogeneity of the subject. However, to be more fruitful, the empirically obtained data ought to be linked with concepts and constructions by means of correspondence rules to create a scientific system. The question is probably not whether research in this area should be theory-driven or practice-driven, but a combination of both that is most fruitful. Fuller (1997), in this context, speaks of "empirically derived theory." Also on the topic of rewarding and road safety, a good starting point for future research would be the empirical findings so far, taking into account the methodological flaws of some studies, and attempt to integrate these with theory. It is, however, beyond the scope of the present study to develop such a theoretical basis for future research in this area.

References

Abelson, R. P. & Levi, A. (1985). Decision making and decision theory. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology. Volume I*, 3rd edn. (pp. 231–309). New York, NY: Random House.

Ajzen, I. (1985). From intentions to actions: A theory of planned behaviour. In J. Kohl & J. Beckman (Eds.), *Action-control: From cognition to behaviour* (pp. 11–39). Heidelberg: Springer-Verlag.

Atkinson, J. W. (1964). *An introduction to motivation*. Princeton, NJ: Van Nostrand Reinhold.

Balch, G. I. (1980). The stick, the carrot, and other strategies. In J. Brigham & D. W. Brown (Eds.), *Policy implementation: penalties or incentives?* (pp. 43–68). London: Sage.

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1986). *Social foundations of thought and action. A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.

Baum, H. & Kling, T. (1997). *Verbesserung der Verkehrssicherheit durch Versicherungsanreize*. BASt Heft M 82, Bergisch Gladbach: Bundesanstalt für Strassenwesen BASt.

Berkowitz, L. (1980). *A survey of social psychology (2nd ed.)*. New York, NY: Holt, Rinehart and Winston.

Bijleveld, F. D. (1998). *Het effect van de bonus/malus-regeling op schademeldingen*. Tech. Rep. R-98-47, Leidschendam: SWOV Institute for Road Safety Research.

Boonstra, J. (1985). Humanisering of beheersing van de arbeid. Een tegenstelling in het streven naar veiligheid. In H. Steensma, R. Van der Vlist & J. Allegro (Eds.), *Modern organiseren en menselijker werken*. Den Haag/Leiden: VUGA, DSWO.

Boonstra, J., Dessens, H., Dijkstra, P., Van de Eshof, W., De Lange, P., Schut, M. & De Vries, H. (1982). *Veiligheid, tot welke prijs?* Tech. rep., Leiden: Vakgroep Sociale en Organisatiepsychologie, Leiden University.

Bower, G. H. (1991). Incentive programmes for promoting safer driving. In M. J. Koornstra & J. Christensen (Eds.), *Enforcement and rewarding: Strategies and effects* (pp. 8–16). Leidschendam: SWOV Institute for Road Safety Research.

Brehm, J. (1966). *A theory of psychological reactance*. New York, NY: Academic Press.

Brehmer, B. (1994). Psychological aspects of traffic safety. *European Journal of Operational Research*, *75*, 540–552.

Brigham, T. A., Meier, S. M. & Goodner, V. (1995). Increasing designated driving with a program of prompts and incentives. *Journal of Applied Behavior Analysis*, *28*, 83–84.

Burke, J. A., Naughton, M. J., Becker, S. L., Arbogast, R., Lauer, R. M. & Krohn, M. D. (1987). The short-term effects of competition and rewards in an adolescent smoking prevention program. *Health Education Quarterly*, *14*, 141–152.

Cairney, P. T. & Catchpole, J. E. (1991). *Road user behaviours which contribute to accidents at urban arterial/local intersections*. Report ARR 197, Australian Road Research Board.

Cameron, J. & Pierce, W. D. (1996). The debate about rewards and intrinsic motivation: Protests and accusations do not alter the results. *Review of Educational Research*, *66*, 39–51.

Campbell, B. J. (1991). Incentives for corporate and political behavior. In M. J. Koornstra & J. Christensen (Eds.), *Enforcement and rewarding: Strategies and effects* (pp. 123–125). Leidschendam: SWOV Institute for Road Safety Research.

Campbell, B. J., Hunter, W. W., Gemming, M. G. & Stewart, J. R. (1984). *The use of economic incentives and public education to increase seat belt use in a community. Final Report*. Tech. Rep. 1984-03, Chapel Hill, NC: University of North Carolina, Highway Safety Research Center.

Campbell, B. J., Hunter, W. W., Stewart, J. R. & Stutts, J. C. (1982). *Increasing safety belt use through an incentive program. Final Report*. Tech. Rep. 1982-10, Chapel Hill, NC: University of North Carolina, Highway Safety Research Center.

Caulkins, M. D. (1971). A municipal safety program that works. *The American City*, *August*, 67–68.

Caverson, R. J. E., Douglas, R. R., Gliksman, L. & Chuipka, L. (1990). Community receptivity to a countermeasure designed to reward sober drivers. *Health Promotion International*, *5*, 119–125.

Condry, J. & Chambers, J. (1978). Intrinsic motivation and the process of learning. In M. R. Lepper & D. Greene (Eds.), *The hidden costs of reward: New perspectives on the psychology of human motivation* (pp. 61–84). Hillsdale, NJ: Lawrence Erlbaum.

Coonley, P. S. & Gurvitz, C. E. (1983). *Assessment of insurance incentives for safety belt usage*. Tech. Rep. 1983-05, Cambridge, MA: Transportation Systems Center.

Cope, J. G., Grossnickle, W. F. & Geller, E. S. (1986a). An evaluation of three corporate strategies for safety belt use promotion. *Accident Analysis & Prevention*, *18*, 243–251.

Cope, J. G., Smith, G. A. & Grossnickle, W. F. (1986b). The effect of variable-rate cash incentives on safety belt use. *Journal of Safety Research*, *17*, 95–99.

Deci, E. L. (1975). *Intrinsic motivation*. New York, NY: Plenum.

Deci, E. L. (1978). Applications of research on the effects of rewards. In M. R. Lepper & D. Greene (Eds.), *The hidden costs of reward: New perspectives on the psychology of human motivation* (pp. 193–203). Hillsdale, NJ: Lawrence Erlbaum.

Deci, E. L. & Porac, J. (1978). Cognitive evaluation theory and the study of human motivation. In M. R. Lepper & D. Greene (Eds.), *The hidden costs of reward: New perspectives on the psychology of human motivation* (pp. 149–176). Hillsdale, NJ: Lawrence Erlbaum.

Deci, E. L. & Ryan, R. M. (1980). The empirical exploration of intrinsic motivational processes. In L. Berkowitz (Ed.), *Advances in experimental social psychology, Vol. 13* (pp. 39–80). New York, NY: Academic Press.

Deci, E. L. & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York, NY: Plenum.

Dickinson, A. M. (1989). The detrimental effects of extrinsic reinforcement on 'intrinsic motivation'. *The Behavior Analyst*, *12*, 1–15.

Dwyer, T. & Raftery, A. E. (1991). Industrial accidents are produced by social relations of work: A sociological theory of industrial accidents. *Applied Ergonomics*, *22*, 167–178.

Dwyer, W. O., Leeming, F. C., Cobern, M. K., Porter, B. E. & Jackson, J. M. (1993). Critical review of behavioral interventions to preserve the environment. Research since 1980. *Environment and Behavior*, *25*, 275–321.

Eagly, A. H. & Chaiken, S. (1993). *The psychology of attitudes*. Orlando, CA: Harcourt Brace.

Eddy, J. M., Marotto, D. A. & Beltz, S. (1988). Analysis of the effectiveness of a safety belt incentive program at CIGNA Corporation. *American Journal of Health Promotion*, *2*, 31–38.

Efron, B. & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York, NY: Chapman & Hall.

Eisenberger, R. & Cameron, J. (1996). Detrimental effects of reward - reality or myth? *American Psychologist*, *51*, 1153–1166.

Elliott, B. (1993). *Road safety mass media campaigns: A meta analysis*. Tech. Rep. CR 118, Canberra: Federal Office of Road Safety FORS.

Elman, D. & Killebrew, T. J. (1978). Incentives and seat belts: Changing a resistant behavior through extrinsic motivation. *Journal of Applied Social Psychology*, *8*, 72–83.

Evans, L. (1987). Belted and unbelted driver accident involvement rates compaired. *Journal of Safety Research*, *18*, 55–67.

Evans, L. (1991). *Traffic safety and the driver*. New York, NY: Van Nostrand Reinhold.

Evans, L. (1996). Safety-belt effectiveness: The influence of crash severity and selective recruitment. *Accident Analysis & Prevention*, *28*, 423–433.

Ferguson, B. & Blampied, N. M. (1991). Unenlightened: An unsuccessful attempt to promote the use of cycle lights at night. *Accident Analysis & Prevention*, *23*, 561–571.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.

Festinger, L. (1964). *Conflict, decision, and dissonance*. Stanford, CA: Stanford University Press.

Fildes, B. (1995). Driver behavior and road safety. In N. Brewer & C. Wilson (Eds.), *Psychology and policing* (pp. 31–61). Hillsdale, NJ: Lawrence Erlbaum.

Foss, R. D. (1989). Evaluation of a community-wide incentive program to promote safety restraint use. *American Journal of Public Health*, *79*, 304–306.

Fox, D. K., Hopkins, B. L. & Anger, W. K. (1987). The long-term effects of a token economy on safety performance in open pit mining. *Journal of Applied Behavior Analysis*, *20*, 215–224.

Friedland, M. L. (Ed.) (1989). *Sanctions and rewards in the legal system: A multidisciplinary approach*. Toronto: University of Toronto Press.

Fuller, R. (1984). A conceptualization of driving behaviour as threat avoidance. *Ergonomics*, *27*, 1139–1155.

Fuller, R. (1991a). Behavior analysis and unsafe driving: Warning-learning trap ahead. *Journal of Applied Behavior Analysis*, *24*, 73–75.

Fuller, R. (1991b). The modification of individual road user behaviour. In M. J. Koornstra & J. Christensen (Eds.), *Enforcement and rewarding: Strategies and effects* (pp. 33–40). Leidschendam: SWOV Institute for Road Safety Research.

Fuller, R. (1997). The psychology of theories and a theory of psychology. In T. Rothengatter & E. Carbonell Vaya (Eds.), *Traffic and transport psychology. Theory and application* (pp. 53–58). Amsterdam/New York: Pergamon.

Geller, E. S. (1982). *Corporate incentives for promoting safety belt use: Rationale, guidelines, and examples*. Tech. Rep. 1982-10-01, Blacksburg, VA: Virginia Polytechnic Institute & State University.

Geller, E. S. (1983a). *Development of industry-based strategies for motivating seat-belt use*. Tech. Rep. DOT HS-806 390, Blacksburg, VA: Virginia Polytechnic Institute & State University.

Geller, E. S. (1983b). Rewarding safety belt usage at an industrial setting: Tests of treatment generality and response maintenance. *Journal of Applied Behavior Analysis*, *16*, 189–202.

Geller, E. S. (1984a). A delayed reward strategy for large-scale motivation of safety belt use: A test of long-term impact. *Accident Analysis & Prevention*, *16*, 457–463.

Geller, E. S. (1984b). *Motivating safety belt use with incentives: A critical review of the past and a look to the future*. SAE Technical Paper 840326, Warrendale: Society of Automotive Engineers.

Geller, E. S. (1988). A behavioral science approach to transportation safety. *Bulletin of the New York Academy of Medicine*, *64*, 632–661.

Geller, E. S. (1989a). Applied behavior analysis and social marketing: An integration for environmental preservation. *Journal of Social Issues*, *45*, 17–36.

Geller, E. S. (1989b). Intervening to increase children's use of safety belts. *Alcohol, Drugs and Driving*, *5*, 37–59.

Geller, E. S. (1990). Where have all the flowers gone? *Journal of Applied Behavior Analysis*, *23*, 269–273.

Geller, E. S. (1992). Applications of behavior analysis to prevent injuries from motor vehicle crashes. In S. S. Glenn (Ed.), *Progress in behavioral studies* (pp. 1–58). Cambridge, MA: Cambridge Center for Behavioral Studies.

Geller, E. S., Berry, T. D., Ludwig, T. D., Evans, R. E., Gilmore, M. R. & Clarke, S. W. (1990). A conceptual framework for developing and evaluating behavior change interventions for injury control. *Health education research: Theory and practice*.

Geller, E. S. & Bigelow, B. E. (1984). Development of corporate incentive programs for motivating safety belt use: A review. *Traffic Safety Evaluation Research Review*, *3*, 21–38.

Geller, E. S., Davis, L. & Spicer, K. (1983). Industry-based incentives for promoting seat belt use: Differential impact on white-collar versus blue-collar employees. *Journal of Organizational Behavior Management*, *5*, 17–29.

Geller, E. S. & Hahn, H. A. (1984). Promoting safety belt use at industrial sites: An effective program for blue collar employees. *Journal of Professional Psychology: Research and Practice*, *15*, 553–564.

Geller, E. S., Johnson, R. P. & Pelton, S. L. (1982a). Community-based interventions for encouraging safety belt use. *American Journal of Community Psychology*, *10*, 183–195.

Geller, E. S., Kalsher, M. J., Rudd, J. R. & Lehman, G. R. (1989). Promoting safety belt use on a university campus: An integration of commitment and incentive strategies. *Journal of Applied Social Psychology*, *19*, 3–19.

Geller, E. S., Paterson, L. & Talbott, E. (1982b). A behavioral analysis of incentive prompts for motivating seat belt use. *Journal of Applied Behavior Analysis*, *15*, 403–413.

Geller, E. S., Rudd, J. R., Kalsher, M. J., Streff, F. M. & Lehman, G. R. (1987). Employer-based programs to motivate safety belt use: A review of short-term and long-term effects. *Journal of Safety Research*, *18*, 1–17.

Geller, E. S., Winett, R. A. & Everett, P. B. (1982c). *Preserving the environment: New strategies for behavior change*. New York, NY: Pergamon.

Gemming, M., Runyan, C., Hunter, W. & Campbell, B. (1984). A community health education approach to occupant protection. *Health Education Quarterly*, *11*, 147–158.

Gifi, A. (1981). *HOMALS user's guide*. Leiden: Department of Datatheory, Leiden University.

Glass, G. V., McGaw, B. & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.

Goldenbeld, C. (1993). *Handhaving van verkeersregels in Nederland; inventarisatie van handhavingsmethodieken. Ervaringen in Nederland*. Tech. Rep. R-93-66, Leidschendam: SWOV Institute for Road Safety Research.

Golob, T. F. & Hensher, D. A. (1994). *Driver behavior of long distance truck drivers: The effects of schedule compliance on drug use and speeding citations*. Tech. Rep. 1994-05, Irvine, CA: California University, Institute of Transportation Studies.

Grant, B. A. (1989). *Differential response of vehicle occupants to seat belt enforcement programs*. Paper presented at the International Symposium on Driving Behaviour in a Social Context (under the patronage of EC/ECMT/WHO and organized by PRI). Paris: 16–18 May 1989.

Grant, B. A., Wilson, R. J. & Dussault, C. (1991). Increasing the use of seat belts through selective traffic enforcement programmes. In M. J. Koornstra & J. Christensen (Eds.), *Enforcement and rewarding: Strategies and effects* (pp. 69–73). Leidschendam: SWOV Institute for Road Safety Research.

Gray, I. (1990). An attempt to reduce accidents in a company car fleet by driver training and encouragement of low risk driving habits. *Journal of Traffic Medicine*, *18*, 139–41.

Grayson, G. B. (1996). *Behavioural adaptation: A review of the literature*. TRL Report 254, Crowthorne: Transport Research Laboratory.

Gregersen, N. P., Brehmer, B. & Morén, B. (1996). Road safety improvement in large companies. An experimental comparison of different measures. *Accident Analysis & Prevention*, *28*, 297–306.

Gregersen, N. P. & Morén, B. (1990). *Traffic safety measures in the Swedish telecommunication administration (Swedish Telecom). A comprehensive study on reducing accidents and costs with driver training, group discussions, bonus and campaign*. VTI Rapport 348, Linköping: Swedish National Road and Transport Research Institue VTI.

Gundy, C. M. (1986). *De effecten van een combinatie van politietoezicht en voorlichting op het gebruik van autogordels*. Tech. Rep. R-86-26, Leidschendam: SWOV Institute for Road Safety Research.

Gundy, C. M. (1988). The effectiveness of a combination of police enforcement and public information for improving seat-belt use. In J. A. Rothengatter & R. A. De Bruin (Eds.), *Road user behavior: Theory and research*. Assen/Maastricht: Van Gorcum.

Hagenzieker, M. P. (1988). *Toepassingen van vormen van individuele beloningen ter bevordering van verkeersveilig gedrag*. Tech. Rep. R-88-12, Leidschendam: SWOV Institute for Road Safety Research.

Hagenzieker, M. P. (1989). *Evaluatie beloningsactie gordelgebruik in Friesland*. Tech. Rep. R-89-20, Leidschendam: SWOV Institute for Road Safety Research.

Hagenzieker, M. P. (1990). *Evaluatie-onderzoek autogordelcampagne onder defensiepersoneel*. Tech. Rep. R-90-14, Leidschendam: SWOV Institute for Road Safety Research.

Hagenzieker, M. P. (1991). Enforcement or incentives? Promoting safety belt use among military personnel in the Netherlands. *Journal of Applied Behavior Analysis*, *24*, 23–30.

Hagenzieker, M. P. (1992a). Beloningen voor verkeersveilig gedrag. *De Psycholoog*, *September*, 349–353.

Hagenzieker, M. P. (1992b). Drivers' opinions of enforcement and incentive strategies to promote safety belt use. *Journal of Safety Research*, *23*, 199–206.

Hagenzieker, M. P., Bijleveld, F. D. & Davidse, R. D. (1997). Effects of incentive programs to stimulate safety belt use: A meta-analysis. *Accident Analysis & Prevention*, *29*, 759–777.

Hale, A. R. (1991). Modifying individual road user behaviour. In M. J. Koornstra & J. Christensen (Eds.), *Enforcement and rewarding: Strategies and effects* (pp. 19–24). Leidschendam: SWOV Institute for Road Safety Research.

Hale, A. R. & Glendon, A. I. (1987). *Individual behaviour in the control of danger*. Amsterdam: Elsevier.

Hall, C. S. & Lindzey, G. (1978). *Theories of personality*. New York, NY: John Wiley & Sons.

Harano, R. M. & Hubert, D. E. (1974). *An evaluation of California's 'Good Driver' incentive program*. Sacramento, CA: California State Department of Motor Vehicles.

Haynes, R., Pine, R. C. & Fitch, H. G. (1982). Reducing accident rates with organizational behavior modification. *Academy of Management Journal*, *25*, 407–416.

Heckhausen, H. (1991). *Motivation and action*. Berlin: Springer-Verlag.

Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, CA: Academic Press.

Heino, A. (1996). *Risk taking in car driving; perceptions, individual differences and effects of safety incentives*. Ph.D. thesis, Groningen: University of Groningen.

Heino, A., Van der Molen, H. H. & Wilde, G. J. S. (1996). Effect van beloningen voor veilig rijden op rijsnelheid en risicoperceptie. In F. J. J. M. Steyvers & P. G. M. Miltenburg (Eds.), *Gedragsbeïnvloeding in verkeers- en vervoersbeleid* (pp. 71–73). Haren: Traffic Research Centre, University of Groningen.

Hensher, D. A., Battellino, H. C., Gee, J. L. & Daniels, R. F. (1991). *Long distance truck drivers on-road performance and economic reward*. Tech. Rep. CR 99, Canberra: Federal Office of Road Safety FORS.

Horne, T. D. & Terry, C. T. (1983a). GM's great seatbelt sweepstakes. *Traffic Safety*, *July/August*, 6/9–28/29.

Horne, T. D. & Terry, C. T. (1983b). *Seat belt sweepstakes – An incentive program*. SAE Technical Paper 830474, Warrendale: Society of Automotive Engineers.

HTS Leeuwarden (1989). *Draagpercentages autogordels; een verslag onderzoeksperiode februari/juli/oktober 1987, maart/mei/ november 1988 in opdracht van Regionaal Orgaan voor de Verkeersveiligheid in Friesland*. Tech. Rep. ROF 89/41/4B, Leeuwarden: HTS Leeuwarden, Afdeling Verkeerstechniek.

Huguenin, R. D. (1997). Do we need traffic psychology models? In T. Rothengatter & E. Carbonell Vaya (Eds.), *Traffic and transport psychology. Theory and application* (pp. 31–39). Amsterdam/New York: Pergamon.

Hunter, J. E. & Schmidt, F. L. (1990). *Methods of meta-analysis. Correcting error and bias in research findings*. Newbury Park, CA: Sage.

Hunter, W. W., Campbell, B. J. & Stewart, J. R. (1986). Seat belts pay off: The evaluation of a community-wide incentive program. *Journal of Safety Research*, *17*, 23–31.

Hunter, W. W., Marchetti, L. M., Stutts, J. C., Little, C. M., Stewart, J. R. & Reinfurt, D. W. (1991a). *An evaluation of non-sanction community seat belt law enforcement programs*. Tech. Rep. HSCR-PR172, Chapel Hill, NC: University of North Carolina, Highway Safety Research Center.

Hunter, W. W., Stewart, J. R., Stutts, J. C., Marchetti, L. M., Little, C. M. & Reinfurt, D. W. (1991b). Non-sanction seat belt law enforcement: A modern day tale of two cities. In *Proceedings of the 35th annual conference of the Association for the Advancement of Automotive Medicine, Toronto, Canada, October 7-9, 1991* (pp. 57–75). Illinois, IL: Association of the Advancement of Automotive Medicine.

Hurst, P. M. (1980). Can anyone reward safe driving? *Accident Analysis & Prevention*, *12*, 217–220.

Janke, M. K. (1990). Safety effects of relaxing California's clean-record requirement for driver license renewal by mail. *Accident Analysis & Prevention*, *22*, 335–349.

Janssen, W. H. (1988). *Risk compensation and the effect of an incentive: A laboratory study*. Report IZF 1988 C-26, Soesterberg: TNO Institute for Perception.

Janssen, W. H. (1989). *An experimental evaluation of safety incentive schemes*. Tech. Rep. IZF 1989 C-19, Soesterberg: TNO Institute for Perception.

Janssen, W. H. (1990). The economy of risk: Or what do I get for my error? *Ergonomics*, *33*, 1333–1348.

Janssen, W. H. (1991). Corporate incentive programmes: An application of utility theory. In M. J. Koornstra & J. Christensen (Eds.), *Enforcement and rewarding: Strategies and effects* (pp. 74–77). Leidschendam: SWOV Institute for Road Safety Research.

Janssen, W. H. (1994). Seat belt wearing and driving behaviour: An empirical investigation. *Accident Analysis & Prevention*, *26*, 249–261.

Janssen, W. H. & Tenkink, E. (1988). Considerations on speed selection and risk homeostasis in driving. *Accident Analysis & Prevention*, *20*, 137–143.

Jeffery, R. W., Forster, J. L., French, S. A., Kelder, S. H., Lando, H. A., McGovern, P. G., Jacobs, D. R. & Baxter, J. E. (1993). The healthy worker project: A work-site intervention for weight control and smoking cessation. *American Journal of Public Health*, *83*, 395–401.

Johnson, R. P. & Geller, E. S. (1984). Contingent versus noncontingent rewards for promoting seatbelt usage. *Journal of Community Psychology*, *12*, 113–122.

Johnston, J. J., Hendricks, S. A. & Fike, J. M. (1994). Effectiveness of behavioral safety belt interventions. *Accident Analysis & Prevention*, *26*, 315–323.

Jonah, B. A. (1997). Sensation seeking and risky driving. In T. Rothengatter & E. Carbonell Vaya (Eds.), *Traffic and transport psychology. Theory and application* (pp. 259–267). Amsterdam/New York: Pergamon.

Jonah, B. A. & Grant, B. A. (1985). Long-term effectiveness of selective traffic enforcement programs for increasing seat belt use. *Journal of Applied Psychology*, *70*, 257–263.

Judd, C. M. & Kenny, D. A. (1981). *Estimating the effects of social interventions*. Cambridge: Cambridge University Press.

Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–291.

Kahneman, D. & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, *39*, 341–350.

Kalsher, M. J., Geller, E. S., Clarke, S. W. & Lehman, G. R. (1989). Safety belt promotion on a naval base: A comparison of incentives vs. disincentives. *Journal of Safety Research*, *20*, 103–113.

Kazdin, A. E. (1984). *Behavior modification in applied settings*. 3rd edn. Homewood, IL: Dorsey.

Kelsey, S. L., Janke, M., Peck, R. C. & Ratz, M. (1985). License extensions for clean-record drivers: A 4-year follow-up. *Journal of Safety Research*, *16*, 149–167.

Kohn, A. (1996). By all available means: Cameron and Pierce's defense of extrinsic motivators. *Review of Educational Research*, *66*, 1–4.

Koornstra, M. J., Mathijssen, M. P. M., Mulder, J. A. G. & Roszbach, R. (Eds.) (1992). *Naar een duurzaam veilig wegverkeer*. Leidschendam: SWOV Institute for Road Safety Research.

KRAFT (1987). *30 Jahre erfolgreiches KRAFT-Fahrsicherheitsprogramm*. KRAFT. Information received from I. H. K. Lameijer.

Kreisverkehrswacht Bautzen (1998). Aktion 'Junge Fahrer'. E-mail message to M. Hagenzieker, 29 June 1998.

Kunreuther, H. (1985). Incentives for improving driving behavior. Ex ante/ex post considerations. In L. Evans & R. C. Schwing (Eds.), *Human behavior and traffic safety* New York, NY: Plenum.

Lehman, G. R. & Geller, E. S. (1990). Educational strategies to increase children's use of safety belts: Are extrinsic rewards necessary? *Health Education Research*, *5*, 187–196.

Lepper, M. R. (1995). Theory by the numbers? Some concerns about meta-analysis as a theoretical tool. *Applied Cognitive Psychology*, *9*, 411–422.

Lepper, M. R. & Greene, D. (Eds.) (1978). *The hidden costs of reward: New perspectives on the psychology of human motivation*. Hillsdale, NJ: Lawrence Erlbaum.

Lepper, M. R., Greene, D. & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the 'overjustification' hypothesis. *Journal of Personality and Social Psychology*, *28*, 129–137.

Lepper, M. R., Keavney, M. & Drake, M. (1996). Intrinsic motivation and extrinsic rewards: A commentary on Cameron and Pierce's meta-analysis. *Review of Educational Research*, *66*, 5–32.

Levelt, P. B. M. (1992). *Beslissingsprocessen van verkeersdeelnemers; covernota bij de rapporten van W.H. Janssen: 'Risk compensation and the effect of an incentive: A laboratory study' (IZF 1988 C-26), 'An experimental evaluation of safety incentive schemes' (IZF 1989 C-19), en 'Seat belt wearing and driving behaviour: An empirical investigation' (IZF 1991 C-15)*. Tech. Rep. R-92-60, Leidschendam: SWOV Institute for Road Safety Research.

Lindzey, G. & Aronson, E. (Eds.) (1985). *The handbook of social psychology. Volume I: Theory and method*. New York, NY: Random House.

Lipsey, M. W. & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, *48*, 1181–1209.

Lonero, L. P., Clinton, K., Wilde, G. J. S., Roach, K., McKnight, A. J., MacLean, H., Guastello, S. J. & Lamble, R. W. (1994). *The roles of legislation, education and reinforcement in changing road user behaviour*. Ontario: Safety Research Office, Safety Policy Branch.

Machemer, E., Runde, B., Wolf, U., Böttner, D. & Tücke, M. (1995). *Delegierte Belohnung und intensivierte Verkehrsüberwachung im Vergleich. Eine empirische Untersuchung zur Beeinflussung des Geschwindigkeitsverhalten*. BASt Heft M 48, Bergisch Gladbach: Bundesanstalt für Strassenwesen BASt.

Mäkinen, T., Wittink, R. D. & Hagenzieker, M. P. (1991). *The use of seat belts and contributing factors - An international comparison*. Tech. Rep. R-91-30, Leidschendam: SWOV Institute for Road Safety Research.

Marsh, W. C. (1978). *Educational approaches to driver improvement*. Tech. Rep. No. 66, Sacramento, CA: California Department of Motor Vehicles.

Mathijssen, M. P. M. (1992). *Rijden onder invloed in de provincie Noord-Brabant 1991/1992. Evaluatie van de alcoholcampagne 1991-1992 van het Regionaal Orgaan voor de Verkeersveiligheid*. Tech. Rep. R-92-25, Leidschendam: SWOV Institute for Road Safety Research.

Mathijssen, M. P. M., Söder, J. C. M. & Verschuur, W. L. G. (1989). Minder automobilisten onder invloed, minder alcoholdoden. *Verkeerskunde*, *40*, 522–525.

Mathijssen, M. P. M. & Verhoef, P. J. G. (1992). *Integraal verkeerstoezicht in de subregio Leiden: Effecten op het gebruik van bromfietshelmen*. Tech. Rep. R-92-18, Leidschendam: SWOV Institute for Road Safety Research.

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.

McCullers, J. C. (1978). Issues in learning and motivation. In M. R. Lepper & D. Greene (Eds.), *The hidden costs of reward: New perspectives on the psychology of human motivation* (pp. 5–17). Hillsdale, NJ: Lawrence Erlbaum.

McGraw, K. O. (1978). The detrimental effects of reward on performance: A literature review and a prediction model. In M. R. Lepper & D. Greene (Eds.), *The hidden costs of reward: New perspectives on the psychology of human motivation* (pp. 31–60). Hillsdale, NJ: Lawrence Erlbaum.

McGuire, W. J. (1985). Attitudes and attitude change. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology. Volume II: Special fields and applications*, 3rd edn. (pp. 233–346). New York, NY: Random House.

McKenna, F. P. (1982). The human factor in driving accidents: An overview of approaches and problems. *Ergonomics*, *25*, 867–877.

McKenna, F. P. (1988). What role should the concept of risk play in theories of accident involvement? *Ergonomics*, *31*, 469–484.

Michon, J. A. (1985). A critical view of driver behaviour models. What do we know, what should we do? In L. Evans & R. Schwing (Eds.), *Human behaviour and traffic safety*. New York, NY: Plenum.

Midden, C. J. H., Westerterp, G. A., Zieverink, H. J. A., Weenig, M. H., Meter, J. E. & Houwen, R. J. (1980). *Energiebesparing, hoe is het mogelijk? Een sociaal-psychologisch onderzoek naar de bevordering van energiebesparing door gedragsbeïnvloeding bij gezinshuishoudingen*. Tech. rep., Petten/Leiden: Energie StudieCentrum ECN/Vakgroep Sociale en Organisatiepsychologie, Leiden University.

Miller, N. & Pollock, V. E. (1995). Use of meta-analysis for testing theory. *Evaluation & the Health Professions*, *18*, 370–392.

Milosevic, S. & Pajevic, D. (1988). Drivers' attitude on seat belt use. In J. A. Rothengatter & R. A. De Bruin (Eds.), *Road user behaviour: Theory and research* (pp. 279–284). Assen/Maastricht: Van Gorcum.

Moore, S. D. (1989). *A meta-analytic review of mass media campaigns designed to change automobile occupant restraint behavior*. Ph.D. thesis, Ann Arbor, MI: University of Illinois at Urbana-Champaign.

Mortimer, R. G. (1991). Motivating seat belt use by incentives and enforcement. In *Proceedings of the Canadian multidisciplinary road safety conference VII, June 17–19, 1991* Vancouver, British Columbia: Department of civil engineering.

Mortimer, R. G., Goldsteen, K., Armstrong, R. W. & Macrina, D. (1990). Effects of incentives and enforcement on the use of seat belts by drivers. *Journal of Safety Research*, *21*, 25–37.

Mulder, J. A. G. & Wesemann, P. (1993). *Beïnvloeding van verkeersgedrag door het prijsmechanisme. Een notitie over mogelijkheden voor verzekeraars*. Tech. Rep. R-93-20, Leidschendam: SWOV Institute for Road Safety Research.

Mullen, B. (1989). *Advanced basic meta-analyses: Procedures and programs*. Hillsdale, NJ: Lawrence Erlbaum.

Murdoch, P. A. & Wilde, G. J. S. (1980). *Reward and punishment in driver control*. Tech. rep., Toronto: Research and Development Division, Ontario Ministry of Transportation and Communication.

Nelson, G. D. & Moffit, P. B. (1988). Safety belt promotion: Theory and practice. *Accident Analysis & Prevention*, *20*, 27–38.

Nimmer, J. G. & Geller, E. S. (1988). Motivating safety belt use at a community hospital: An effective integration of incentive and commitment strategies. *American Journal of Community Psychology*, *16*, 381–395.

OECD (1990). *Behavioural adaptations to changes in the road transport system*. Paris: OECD.

Olson, J. M. & Zanna, M. P. (1993). Attitudes and attitude change. *Annual Review of Psychology*, *44*, 117–154.

Orr, L. (1982). Incentives and efficiency in automobile safety regulation. *Quarterly Review of Economics and Business*, *22*, 43–65.

Oskamp, S. (1984). *Applied social psychology*. New York, NY: Prentice-Hall.

Overend, R. B. (1984). DWI program incentive awards highlight highway traffic safety division meeting. *Traffic Safety*, (pp. 14–16).

Parker, D., Manstead, A. S. R., Stradling, S. G. & Reason, J. T. (1992). Determinants of intention to commit driving violations. *Accident Analysis & Prevention*, *24*, 117–131.

Parker, D., West, R., Stradling, S. G. & Manstead, A. S. R. (1995). Behavioural characteristics and involvement in different types of traffic accident. *Accident Analysis & Prevention*, *27*, 571–581.

Peck, R. C., Kelsey, S. L., Ratz, M. & Sherman, B. R. (1980). The effectiveness of accredited traffic violator schools in reducing accidents and violations. *Journal of Safety Research*, *12*, 68–77.

Perkins, W. (1990). Changing driver behaviour: An overview. In *Road safety: A community challenge. Issues arising from seminars in Wellington, Auckland and Christchurch* (pp. 131–139). New Zealand.

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

Reason, J. T. (1979). Actions not planned: The price of automatisation. In G. Underwood & R. Stevens (Eds.), *Aspects of consciousness 1*. London: Academic Press.

Reason, J. T., Manstead, A. S. R., Stradling, S. G.and Parker, D. & Baxter, J. S. (1991). *The social and cognitive determinants of aberrant driving behaviour*. Tech. Rep. 253, Crowthorne: Transport Research Laboratory.

Roberts, M. C., Alexander, K. & Knapp, L. G. (1990). Motivating children to use safety belts: A program combining rewards and 'Flash for Life'. *Journal of Community Psychology*, *18*, 110–119.

Roberts, M. C. & Fanurik, D. (1986). Rewarding elementary schoolchildren for their use of safety belts. *Health Psychology*, *5*, 185–196.

Roberts, M. C., Fanurik, D. & Wilson, D. R. (1988). A community program to reward children's use of seat belts. *American Journal of Community Psychology*, *16*, 395–407.

Roberts, M. C. & Turner, D. S. (1986). Rewarding parents for their children's use of safety seats. *Journal of Pediatric Psychology*, *11*, 25–36.

Robertson, L. S. (1984). Insurance incentives and seat belt use. *American Journal of Public Health*, *74*, 1157–1158.

Rooijers, A. J. & De Bruin, R. A. (1991). Selective enforcement of speeding behaviour in built-up areas. In M. J. Koornstra & J. Christensen (Eds.), *Enforcement and rewarding: Strategies and effects* (pp. 51–54). Leidschendam: SWOV Institute for Road Safety Research.

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin*, *118*, 183–192.

Rothengatter, J. A., De Waard, D., Slotegraaf, J., Carbonell Vaya, E. & Muskaug, R. (1992). *Social acceptance of automatic policing and information systems*. DRIVE project V1033 Autopolis. Groningen: Traffic Research Centre, University of Groningen.

Rothengatter, T. (1997). Errors and violations as factors in accident causation. In T. Rothengatter & E. Carbonell Vaya (Eds.), *Traffic and transport psychology. Theory and application* (pp. 59–64). Amsterdam/New York: Pergamon.

Rothengatter, T. & Carbonell Vaya, E. (Eds.) (1997). *Traffic and transport psychology. Theory and application*. Amsterdam/New York: Pergamon.

Rudd, J. R. & Geller, E. S. (1985). A university-based incentive program to increase safety belt use: Toward cost-effective institutionalization. *Journal of Applied Behavior Analysis*, *18*, 215–226.

Rumar, K. (1988). Collective risk but individual safety. *Ergonomics*, *31*, 507–518.

Ryan, R. M. & Deci, E. L. (1996). When paradigms clash: Comments on Cameron and Pierce's claim that rewards do not undermine intrinsic motivation. *Review of Educational Research*, *66*, 33–38.

Sabey, B. E. & Taylor, H. (1980). *The known risks we run: The highway*. Report 567, Crowthorne: Transport Research Laboratory.

Sadler, D. D., Perrine, M. W. & Peck, R. C. (1991). The long-term traffic safety impact of a pilot alcohol abuse treatment as an alternative to license suspensions. *Accident Analysis & Prevention*, *23*, 203–224.

Sagberg, F. (1994). *Motivasjon og trafikksikkerhet blant yrkessjaafoerer; utproeving av skadeforebyggende tiltak ved Oslo postterminal*. TØI rapport 263, Oslo, Norway: Transportoekonomisk Institutt TØI.

Sagberg, F., Fosser, S. & Sætermo, I. A. F. (1997). An investigation of behavioural adaptation to airbags and antilock brakes among taxi drivers. *Accident Analysis & Prevention*, *29*, 293–302.

Sarafino, E. P. (1984). Rewards and intrinsic interest. In R. J. Corsini (Ed.), *Encyclopedia of psychology*, vol. 3 (pp. 238–239). New York, NY: John Wiley & Sons.

SAS Institute (1985). *SAS user's guide: Statistics, Version 5 Edition*. Cary, NC: SAS Institute Inc.

SAS Institute (1990). *SAS/STAT users' guide, Version 6*, vol. 2. 4th edn. Cary, NC: SAS Institute Inc.

Schaaf, W. & Granderath, F. J. (1989). Young drivers' campaign in Germany. In *Proceedings of the International conference on new ways and means for improved safety, Tel Aviv, Israel, February 20–23, 1989*. Haifa: Transportation Research Institute Technion.

Schmidt, T. (1989). Straf en beloning als instrumenten om de modal split te beïnvloeden; kilometerreductie-contracten bieden mogelijk praktische aanknopingspunten. *Verkeerskunde*, *40*, 332–335.

Shadish, W. R., Cook, T. D. & Levinton, L. C. (1991). *Foundations of program evaluation. Theories of practice*. Newbury Park, CA: Sage.

Simmelink, I. & Heere, F. (1993). *Met de waakvlam in de pijp. Optimalisering van veilig en goedkoop rijgedrag van vrachtwagenchauffeurs*. Tech. Rep. 1058, Leiden: Research voor Beleid.

Simons-Morton, B. G., Brink, S. & Bates, D. (1987). Effectiveness and cost effectiveness of persuasive communications and incentives in increasing safety belt use. *Health Education Quarterly*, *14*, 167–179.

Skinner, B. F. (1938). *The behavior of organisms*. New York, NY: Appleton–Century–Croft.

Skinner, B. F. (1953). *Science and human behavior*. New York, NY: Macmillan.

Slovic, P., Fischhoff, B. & Lichtenstein, S. (1982). Facts versus fears: Understanding percieved risks. In P. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

Spoonhour, K. A. (1981). Company 'Snap it up' campaign achieves 90 percent belt use. *Traffic Safety*, (pp. 18–20).

Stevens, M. M., Paine-Andrews, A. & Franciso, V. T. (1996). Improving employee health and wellness: A pilot study of the employee-driven Perfect Health Program. *American Journal of Health Promotion*, *11*, 12–14.

Strecher, V. J., Seijts, G. H., Kok, G. J., Latham, G. P., Glasgow, R., DeVellis, B., Meertens, R. M. & Bulger, D. W. (1995). Goal setting as a strategy for health behavior change. *Health Education Quarterly*, *22*, 190–200.

Streff, F. M. & Geller, E. S. (1986). Strategies for motivating safety belt use: The application of applied behavior analysis. *Health Education Research*, *1*, 47–59.

Streff, F. M. & Wagenaar, A. C. (1989). Are there really shortcuts? Estimating seat belt use with self-report measures. *Accident Analysis & Prevention*, *21*, 509–516.

Stutts, J. C., Hunter, W. W. & Campbell, B. J. (1984). Three studies evaluating the effectiveness of incentives for increasing safety belt use. *Traffic Safety Evaluation Research Review*, *3*, 9–20.

Thyer, B. A. & Geller, E. S. (1990). Behavior analysis in the promotion of safety-belt use: A review. In M. Hersen, R. M. Eisler & P. Miller (Eds.), *Progress in behavior modification*, vol. 26 (pp. 150–172). Newbury Park, CA: Sage.

Tränkle, U. & Gelau, C. (1992). Maximization of subjective expected utility or risk control? Experimental test of risk homeostasis theory. *Ergonomics*, *35*, 7–23.

Tücke, M. & Wolf, U. (1996). Reduzierung von Geschwindigkeiten im ländlichen Raum durch 'delegierte Belohnung' und/oder intensivierte polizeiliche überwachung. In *Conference folder of the fifth European workshop on recent developments in road safety research, May 3rd, 1996, Bern, Switzerland; Topic 4: Measures to increase traffic law compliance*

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.

Twisk, D. A. M. (1990). *De verkeersveiligheid van jonge, onervaren automobilisten en de invoering van een voorlopig rijbewijs; Een literatuurstudie*. Tech. Rep. R-90-44, Leidschendam: SWOV Institute for Road Safety Research.

Twisk, D. A. M. & Mulder, J. A. G. (1992). *Verzekeringsmaatschappijen en verkeersongevallen*. Tech. Rep. R-92-38, Leidschendam: SWOV Institute for Road Safety Research.

Ullmann, L. P. (1984). Behavior modification. In R. J. Corsini (Ed.), *Encyclopedia of psychology*, vol. 1 (pp. 134–137). New York, NY: John Wiley & Sons.

Underwood, G., Chapman, P., Wright, S. & Crundall, D. (1997). Estimating Accident Liability. In T. Rothengatter & E. Carbonell Vaya (Eds.), *Traffic and transport psychology. Theory and application* (pp. 247–258). Amsterdam/New York: Pergamon.

Vaaje, T. (1991). Rewarding in insurance: Return of part of premium after a claim-free period. In M. J. Koornstra & J. Christensen (Eds.), *Enforcement and rewarding: Strategies and effects* (pp. 154–156). Leidschendam: SWOV Institute for Road Safety Research.

Van Houten, R. & Malenfant, J. E. L. (1993). A multifaceted approach to reduce risk taking behavior by road users. In *University of Michigan symposium on Risk Taking Behavior and Traffic Safety* (pp. 60–63).

Van Houten, R. & Nau, P. A. (1983). Feedback interventions and driving speed: A parametric and comparative analysis. *Journal of Applied Behavior Analysis*, *16*, 253–288.

Van Houten, R., Rolider, A., Nau, P. A., Friedman, R., Becker, M., Chalodovsky, I. & Scherer, M. (1985). Large-scale reductions in speeding and accidents in Canada and Israel: A behavioural ecological perspective. *Journal of Applied Behavior Analysis*, *18*, 87–93.

Van Kampen, L. T. B. (1988). *Analyse van de verkeersonveiligheid van jonge, onervaren automobilisten; Een probleemanalyse*. Report R-88-45, Leidschendam: SWOV Institute for Road Safety Research.

Van Olst, E. H. & Bakker, S. J. (1979). Conditioning. In J. A. Michon, E. G. J. Eijkman & L. F. W. De Klerk (Eds.), *Handbook of psychonomics*, vol. 2 (pp. 23–74). Amsterdam: North-Holland.

Varkevisser, G. A. & Arnoldus, J. G. (1989). *Aanwezigheid en gebruik van autogordels 1988*. Tech. Rep. R-89-11, Leidschendam: SWOV Institute for Road Safety Research.

Verhoef, P. J. G. (1991). *Aanwezigheid en gebruik van autogordels 1990*. Tech. Rep. R-91-6, Leidschendam: SWOV Institute for Road Safety Research.

Vlek, C., Jager, W. & Steg, L. (1997). Modellen en strategieën voor gedragsverandering ter beheersing van collectieve risico's. *Nederlands Tijdschrift voor de Psychologie*, *52*, 174–191.

Vlek, C. & Wagenaar, W. A. (1979). Judgement and decision under uncertainty. In J. A. Michon, E. G. J. Eijkman & L. F. W. De Klerk (Eds.), *Handbook of psychonomics*, vol. 2 (pp. 253–345). Amsterdam: North-Holland.

Wagenaar, W. A. (1990). Risk evaluation and the causes of accidents. In K. Borcherding, O. I. Larichev & D. M. Messick (Eds.), *Contemporary issues in decision making*. Amsterdam: North-Holland.

Warner, K. & Murt, H. (1984). Economic incentives for health. *Annual Review of Public Health*, *5*, 107–133.

Weiner, B. (1992). *Human motivation. Metaphors, theories, and research*. Newbury Park, CA: Sage.

Weinstein, N., Grubb, P. & Vautier, J. (1986). Increasing automobile seat belt use: An intervention emphasizing risk susceptibility. *Journal of Applied Psychology*, *71*, 285–290.

Wilde, G. J. S. (1985). The use of incentives for the promotion of accident-free driving. *Journal of Studies on Alcohol*, (pp. 161–167).

Wilde, G. J. S. (1988). Incentives for safe driving and insurance management. An evaluation of the camparative feasibility and effectiveness of incentive programmes for safe driving in the general population under conditions of publicly- versus privately-administered automobile insurance. In C. A. Osborne (Ed.), *Report of inquiry into motor vehicle accident compensation in Ontario* (pp. 464–511). Toronto: Ontario Ministry of the Attorney General and the Ministry of Financial Insititutions.

Wilde, G. J. S. (1994). *Target risk*. PDE Publications.

Wilde, G. J. S. & Murdoch, P. A. (1982). Incentive systems for accident-free and violation-free driving in the general population. *Ergonomics*, *25*, 879–890.

Wilde, G. J. S., O'Neill, B. & Cannon, D. (1975). *A psychometric investigation of drivers' concern for road safety and their opinions of various measures for accident prevention*. Studies of safety in transport, Kingston, Ontario: Queen's University.

Williams, A. F., Lund, A. K., Preusser, D. F. & Blomberg, R. D. (1987a). Results of a seat belt use law enforcement and publicity campaign in Elmira, New York. *Accident Analysis & Prevention*, *19*, 243–249.

Williams, A. F., Wells, J. K. & Lund, A. K. (1987b). Shoulder belt use in four states with belt use laws. *Accident Analysis & Prevention*, *19*, 251–260.

Williams, B. W. (1980). Reinforcement, behavior constraint, and the overjustification effect. *Journal of Personality and Social Psychology*, *39*, 599–614.

Winett, R. A., King, A. C. & Altmar, D. G. (1989). Incentives in health promotion: A theoretical framework and applications. In *Health psychology and public health* (pp. 71–92). New York, NY: Pergamon.

Wojtowicz, G. G., Peveler, L. A., Eddy, J. M., Waggle, S. B. & Fitzhugh, E. C. (1992). The midfield high school safety belt incentive program. *Journal of School Health*, *62*, 407–410.

Wolf, U. & Machemer, E. (1994). Wissenschaft und Praxis im Dialog: Möglichkeiten und Bedeutung einer gemeinsamen Verkehrssicherheitsforschung am Beispiel des Projektes Lingen (Ems). In *Internationalen Seminar der Polizei-Fuehrungsakademie* (pp. 227–238). Münster: Polizei-Fuehrungsakademie.

Zaal, D. (1994). *Traffic law enforcement: A review of the literature*. Tech. Rep. 53, Canberra: Monash University, Accident Research Centre.

Zuckerman, M. (1992). *Biological bases of sensation seeking, impulsivity, and anxiety*. Hillsdale, NJ: Lawrence Erlbaum.

Appendix

References for studies in meta-analysis

(1) Campbell et al. (1982)
(2) Campbell et al. (1984)
(3) Cope et al. (1986a)
(4) Cope et al. (1986b)
(5) Eddy et al. (1988)
(6) Elman & Killebrew (1978)
(7) Foss (1989)
(8) Geller (1982)
(9) Geller (1983a)
(10) Geller (1983b)
(11) Geller (1984a)
(12) Geller (1989b)
(13) Geller et al. (1982a)
(14) Geller et al. (1982b)
(15) Geller et al. (1983)
(16) Geller et al. (1989)
(17) Hagenzieker (1989)
(18) Hagenzieker (1991)
(19) Horne & Terry (1983a)
(20) Hunter et al. (1991b)
(21) Johnson & Geller (1984)
(22) Kalsher et al. (1989)
(23) Lehman & Geller (1990)
(24) Mortimer (1991)
(25) Nimmer & Geller (1988)
(26) Roberts & Fanurik (1986)
(27) Roberts & Turner (1986)
(28) Roberts et al. (1988)
(29) Roberts et al. (1990)
(30) Robertson (1984)
(31) Rudd & Geller (1985)
(32) Simons-Morton et al. (1987)
(33) Weinstein et al. (1986)
(34) Wojtowicz et al. (1992).

Samenvatting

Belonen en straffen: het zijn bekende methoden om gedrag te beïnvloeden. In de verkeerspraktijk is het traditioneel zo dat gewenst gedrag 'gestimuleerd' wordt door regels en wetten, zodat overtredingen kunnen worden bestraft. Aan beloningen voor het bevorderen van gewenst verkeersgedrag is tot voor kort weinig aandacht besteed. Het doel van dit proefschrift was om meer inzicht te verkrijgen in de mogelijke rol van beloningen bij het beïnvloeden van verkeers(veilig) gedrag door te onderzoeken in hoeverre beloningen effectief kunnen zijn om verkeersgedrag te veranderen, welke vormen van beloningen daarin meer en minder effectief zijn dan andere, welke factoren en omstandigheden hierop van invloed zijn en na te gaan of deze manier van gedragsbeïnvloeding in de verkeerspraktijk toepasbaar is.

In Hoofdstuk 1 wordt verslag gedaan van een literatuurstudie waarin aandacht wordt besteed aan de rol van beloningen in psychologische theorieën en modellen van verkeersgedrag. Ook is geïnventariseerd wat er in diverse toepassingsgebieden bekend is – blijkend uit evaluatiestudies – over het effect van beloningen op gedrag. Veel psychologische theorieën – meestal op het gebied van leren en motivatie – ruimen een belangrijke plaats in voor 'belonen'. In theorieën variërend van die van Skinner en moderne leertheorieën zoals die van Bandura, tot meer cognitief georiënteerde sociaal-psychologische theorieën zoals Festingers cognitieve-dissonantietheorie en (subjectieve) utiliteitstheorieën, worden beloningen expliciet als krachtig gedragsbeïnvloedend beschouwd; dit hoewel hun uitgangspunten en veronderstelde verklaringsmechanismen heel verschillend zijn. Terwijl leertheorieën een hoofdrol toekennen aan beloningen om gedrag te veranderen, benadrukken de sociaal-cognitief georiënteerde theorieën de interactie tussen beloningen en cognitieve processen zoals (intrinsieke) motivatie. Factoren die blijkens de verschillende theoretische benaderingen van invloed zijn op het effect van beloningen op gedrag zijn bijvoorbeeld de grootte van en het type beloning, de kans op en het tijdstip van het uitreiken van een beloning, het type taak of soort gedrag waarvoor beloond wordt, de sociale 'setting', kosten-baten overwegingen en andere cognitieve evaluaties van deze factoren.

In modellen van verkeersgedrag wordt veel minder vaak aandacht besteed aan de rol van beloningen. De risicohomeostase-theorie van Wilde, een op utiliteitstheorie gebaseerd model van Janssen en het 'threat-avoidance' model van Fuller besteden aandacht aan het effect van beloningen op verkeersgedrag en de verkeersveiligheid. De eerste twee modellen veronderstellen onder andere een positief effect van het in het vooruitzicht stellen van 'uitkomst georiënteerde' beloningen: beloningen voor ongevallenvrij rijden; het model van Fuller veronderstelt onder andere dat externe beloningen ervoor kunnen zorgen dat gevaarlijke verkeersgedragingen die als prettig ervaren worden door bestuurders (bijvoorbeeld hard rijden) omgebogen worden naar de veilige kant.

In verschillende toepassingsgebieden zijn beloningen met succes gebruikt om gewenst gedrag te bevorderen, bijvoorbeeld om werknemers in de industrie er toe te brengen zich aan de veiligheidsvoorschriften te houden, om milieuvriendelijk gedrag te stimuleren en om rook- en eetgedrag te beïnvloeden. Op het gebied van verkeer worden ook positieve effecten van beloningsprogramma's gemeld in de literatuur. Dergelijke programma's zijn vooral toegepast in bedrijven. Bedrijven lijken een goede omgeving voor het toepassen van beloningsacties, omdat het voor hen kostenbesparend kan werken. Door veiliger gedrag op de weg van hun werknemers vermindert bijvoorbeeld het ziekteverzuim, en kan (indirect) de produktiviteit stijgen omdat er minder ongevallen plaatsvinden. Bij deze toepassingen is gewoonlijk sprake van een specifieke, min of meer homogene populatie en een veelal kleinschalige opzet, wat de praktische uitvoering vergemakkelijkt. Ook zijn enige voorbeelden voorhanden van wat grootschaliger beloningsprogramma's voor schadevrij rijden. Het beschikbare onderzoek naar beloningen voor specifiek verkeersveilig gedrag is met name gericht op het dragen van autogordels. Dergelijke beloningsprogramma's hebben over het algemeen geleid tot substantiële verbeteringen in het gordelgebruik. Het probleem is echter dat de meeste studies zijn verricht in situaties waarin gordelgebruik nog vrijwillig was en (derhalve) met een zeer laag gordelgebruik, en dat kenmerken van beloningsprogramma's niet systematisch gevarieerd zijn. Hierdoor is het bijvoorbeeld niet goed mogelijk vast te stellen hoe een 'optimaal' beloningsprogramma er uit moet zien en of beloningsprogramma's ook effectief zijn onder omstandigheden met een relatief hoog beginniveau van gordelgebruik.

Om meer inzicht te verkrijgen in de factoren die de effectiviteit van beloningen beïnvloeden is besloten nader onderzoek te richten op één specifiek gedrag – autogordelgebruik. Een aantal veldstudies is uitgevoerd waarin het effect van verschillende beloningsacties op het dragen van autogordels is onderzocht in situaties met een relatief hoog uitgangspercentage gordeldragers (Hoofdstuk 2–4). Daarbij is ook onderzocht wat het effect van belonen is in vergelijking tot het effect van politietoezicht. Vervolgens is met behulp van een meta-analyse nagegaan hoe de resultaten van deze onderzoeken zich verhouden tot die verkregen uit eerder uitge-

voerd onderzoek naar de effecten van beloningsprogramma's op het gebruik van autogordels, en welke kenmerken van beloningsprogramma's aangewezen kunnen worden die bepalen of een programma meer of minder effectief is (Hoofdstuk 5).

In Hoofdstuk 2 wordt de evaluatie van een kleinschalige beloningsactie beschreven die werd uitgevoerd in de provincie Friesland. Uit deze 'pilot'-studie bleek dat er na afloop van de actie geen verbetering in het gordelgebruik was opgetreden. In tegenstelling tot wat over het algemeen in de literatuur gerapporteerd wordt, werd zelfs geen korte-termijneffect geconstateerd. Mogelijke verklaringen die aangedragen worden voor het uitblijven van enig effect zijn dat de actie niet voldoende bekend was bij het publiek, dat het eenmalig belonen en gedurende slechts één dag niet voldoende is om een gedragsverandering tot stand te brengen en dat de beloning niet als 'aantrekkelijk' beschouwd werd.

Als vervolg op deze pilot studie is een grootschaliger veldstudie uitgevoerd op twaalf Nederlandse kazernes (Hoofdstuk 3 en 4). Diverse campagnevarianten gericht op het stimuleren van het gordelgebruik door personeel in dienst van het Ministerie van Defensie in hun eigen personenauto's zijn onderzocht. Het type beloningsprogramma, type publiciteit en hoeveelheid politietoezicht werd gevarieerd. Om tegemoet te komen aan de beperkingen van de pilot-studie werd veel aandacht besteed aan de publiciteit rondom de campagne, is geprobeerd voor de doelgroep van deze campagne aantrekkelijke beloningen in het vooruitzicht te stellen (o.a. cadeaubonnen, compact disc spelers), en had de campagne een langere duur (twee maanden). Onderscheid werd gemaakt naar verschillende typen beloningsprogramma's. Op twee kazernes werd een wedstrijd georganiseerd tijdens de campagne. Een prijs van *f* 5.000,- kon worden gewonnen door het personeel van die kazerne met het hoogste gordeldraagpercentage aan het eind van de actie (groepsbelonen). Op twee andere kazernes werden tijdens de actie loten uitgereikt aan bestuurders (en voorpassagiers) als zij hun gordel droegen. Eén of meer prijzen werd(en) elke week verloot onder degenen in het bezit van een lot (individueel belonen; variatie in frequentie van belonen). Op de andere kazernes werden verschillende combinaties van hoeveelheid politietoezicht en publiciteit toegepast.

Het bleek dat politietoezicht en beloningsacties gemiddeld even effectief waren: uit gedragsobservaties bleek dat de stijging in gordelgebruik ongeveer 11–14 percentagepunten bedroeg, gemeten over een periode van vlak voor de campagne tot drie maanden na afloop van de campagne. Dus zelfs als gordelgebruik verplicht is en het gebruik van gordels al relatief groot is (aanvangspercentage van ongeveer 65%) blijken beloningen effectief te kunnen zijn. Dit gold vooral voor de individuele beloningsacties. De frequentie van belonen was niet van invloed op de grootte van het effect: het uitreiken van één of vier prijzen per week bleek even effectief. De variant met groepsbelonen en competitie was minder effectief.

Naast gedragsobservaties is de campagne ook geëvalueerd door middel van een schriftelijke enquête onder het personeel van de kazernes (Hoofdstuk 4). De vragen betroffen o.a. beweerd gordelgebruik, motivatie om al dan niet een gordel te dragen, de bekendheid van de campagne en meningen over politietoezicht en beloningsprogramma's. In tegenstelling tot wat uit andere onderzoeken was gebleken, rangschikt deze groep respondenten beloningsacties niet onder de meest favoriete verkeersveiligheidsmaatregelen. Straf als maatregel om gordelgebruik te bevorderen vond meer steun bij de ondervraagde militairen. Overigens kan worden aangenomen dat de mening van deze specifieke groep wellicht niet representatief is voor 'de Nederlandse bevolking'. Het is wel zo dat respondenten die te maken hadden gehad met een van de beloningsacties positiever dachten over 'beloningen' dan respondenten van kazernes waar toezichtacties waren gehouden. Het lijkt er overigens op dat vooral degenen die af en toe een autogordel dragen worden beïnvloed door een beloningsactie, die hen als het ware een extra aanzet geeft tot het vaker dragen van de gordel. Degenen die nooit gordels (zeggen te) dragen lijken nauwelijks te worden gestimuleerd door beloningen. De notie dat externe, materiële beloningen een negatief effect kunnen hebben op de (intrinsieke) motivatie om een gordel te dragen lijkt niet te worden ondersteund. De meest gerapporteerde motieven om een gordel om te doen zijn: vanwege de beschermende eigenschappen van de autogordel in geval van een botsing, uit gewoonte, en omdat het verplicht is. Slechts een kleine minderheid noemt een beloning als motief om de gordel om te doen.

In Hoofdstuk 5 worden de resultaten van een meta-analyse beschreven die is uitgevoerd om te bepalen hoe groot de korte- en lange-termijn effecten zijn van beloningsprogramma's om het gebruik van autogordels te bevorderen. De meta-analyse had tevens tot doel om variabelen te identificeren die van invloed zijn op de grootte van deze effecten. De resultaten van deze meta-analyse bevestigen conclusies uit eerdere, meer 'traditionele' literatuurstudies, dat beloningscampagnes leiden tot aanzienlijke verbeteringen in gordelgebruik op de korte termijn; en dat lange(re)-termijneffecten over het algemeen kleiner zijn dan korte-termijneffecten, maar het gordelgebruik dan nog altijd groter is dan voor aanvang van de beloningscampagne. Terwijl eerdere literatuurstudies deze effecten niet kwantificeren, blijkt uit de resultaten van deze meta-analyse dat de gemiddelde stijging in gordelgebruik op de korte termijn ongeveer 12 percentagepunten (gewogen gemiddelden) bedraagt ten opzichte van het niveau voorafgaand aan de campagne, en op meer lange termijn (meestal enige maanden) ongeveer 10 percentagepunten (gewogen gemiddelden; ongewogen gemiddelden zijn respectievelijk 21 en 14 percentagepunten voor de korte en lange termijn).

De bevindingen van de meta-analyse laten ook zien dat de grootte van de effecten afhankelijk is van een aantal interveniërende variabelen: niet alle beloningsprogramma's zijn even effectief. Het korte-termijneffect wordt vooral bepaald door

(een combinatie van) de volgende variabelen: de doelgroep van de actie, het tijdstip waarop beloningen worden uitgereikt en het beginniveau van het gordelgebruik. Kleinschalige beloningscampagnes onder min of meer homogene groepen (zoals bij werknemers van bedrijven of leerlingen en hun ouders op scholen) leiden tot betere resultaten dan grootschaliger campagnes (zoals die waar alle automobilisten in een bepaalde regio de doelgroep vormen). Het onmiddellijk uitreiken van beloningen leidt tot grotere effecten dan uitgesteld belonen (zoals bij loterijen het geval is); een combinatie van direct en uitgesteld belonen lijkt het meest effectief. Verder blijken beloningsprogramma's tot grotere effecten te leiden naarmate het initiële gordelgebruik lager is (hetgeen hoog correleert met de afwezigheid van de verplichting om een gordel te dragen).

Ook bleek dat groepsbelonen over het algemeen tot grotere effecten leidt dan individuele beloningen, maar deze variabele bleek minder stabiel dan de hiervoor genoemde. Verder bleek dat ook campagnes met een korte duur, met een kleine kans op een beloning en met kleine prijzen effectief zijn; deze variabelen waren niet gerelateerd aan de grootte van het effect. Lange-termijneffecten waren veel lastiger te relateren aan kenmerken van de campagnes. Tenslotte wordt in dit hoofdstuk aandacht besteed aan een aantal praktische en methodologische problemen die zich voordeden bij het uitvoeren van de meta-analyse.

In Hoofdstuk 6 wordt een samenvatting gegeven van de belangrijkste resultaten van het onderzoek dat gerapporteerd wordt in dit proefschrift en worden conclusies getrokken met betrekking tot de effectiviteit van beloningsprogramma's en de toepasbaarheid van beloningen in de verkeerspraktijk. Uit het empirische onderzoek blijkt dat beloningsprogramma's zeer effectief kunnen zijn om gewenst verkeersgedrag te bevorderen, met name worden substantiële korte-termijneffecten gerapporteerd. Dit geldt voor het bevorderen van vrijwillige gedragingen, maar ook voor het stimuleren van verplicht gedrag, zoals gebleken is uit studies op het gebied van het dragen van autogordels. De effecten van beloningen zijn van dezelfde orde van grootte als die van politietoezicht. Ook zogenaamde 'uitkomst-georiënteerde' beloningsprogramma's blijken effectief, deze brengen over het algemeen een reductie in ongevallen met zich mee. Voor dergelijke beloningsprogramma's is het echter nog onduidelijk welke gedragingen tot de ongevallenreductie hebben geleid.

Een aantal voordelen kan worden genoemd van het gebruik van beloningen vergeleken met andere maatregelen, zoals politietoezicht of voorlichtingscampagnes. Het gebruik van beloningen kan bijvoorbeeld kostenbesparend zijn en wordt door het publiek over het algemeen positief gewaardeerd. Aan de andere kant lijkt het niet eenvoudig om beloningen op grote schaal structureel te implementeren. Mede gezien de substantiële effecten die gevonden zijn, wordt het de moeite waard geacht de mogelijkheden voor grootschalige en structurele toepassingen van beloningen in het verkeer nader te bestuderen. Verder zijn er nog veel openstaande vragen

met betrekking tot de effectiviteit van beloningsprogramma's voor andere gedragingen dan het gebruik van autogordels en is door een gebrek aan systematische variatie in de kenmerken van beloningsprogramma's nog onduidelijk welke vorm van belonen het meest effectief is. Op grond van de huidige kennis – ontleend aan empirische resultaten en theoretische overwegingen – wordt een aantal aanbevelingen en richtlijnen gegeven voor het opzetten van beloningsprogramma's in het verkeer.

Tenslotte wordt kort ingegaan op praktijk- en theoretisch 'gestuurd' onderzoek. Het onderzoek op het gebied van beloningen in het verkeer kan voornamelijk als 'gestuurd door de praktijk' worden bestempeld; beloningsacties worden vaak op een ad hoc basis uitgeprobeerd als een mogelijke oplossing voor een specifiek probleem. Hoewel het onderzoek naar de effecten van beloningen op verkeersgedrag wel geïnspireerd lijkt te zijn door theoretische noties, is het over het algemeen niet gericht op het toetsen van specifieke hypothesen die zijn afgeleid uit (een bepaalde) theorie: het is niet 'gestuurd' door theorie. Aanbevolen wordt om in toekomstig onderzoek praktijk- en theoretisch onderzoek meer te combineren en te integreren dan tot nog toe het geval is; verondersteld wordt dat dit tot vruchtbaardere resultaten zal leiden.

Nawoord

Veel mensen hebben mij geholpen bij het tot stand komen van dit proefschrift. Collega's, vrienden en familie hebben de afgelopen jaren hun belangstelling en steun getoond. Ik wil hen allen hiervoor bedanken. De SWOV wil ik danken voor het beschikbaar stellen van tijd en faciliteiten. Een deel van het onderzoek waarover gerapporteerd wordt in dit proefschrift werd financieel ondersteund door de (toenmalige) Dienst Verkeerskunde van het Ministerie van Verkeer en Waterstaat en het Regionaal Orgaan voor de Verkeersveiligheid in Friesland. In het bijzonder wil ik Frits Bijleveld bedanken voor zijn hulp bij het uitvoeren van statistische analyses, voor het verzorgen van de vormgeving van het boek en voor alle andere steun die hij gaf. Piet Noordzij wil ik danken voor de manier waarop hij mij door de jaren heen het vertrouwen gaf dat dit project de moeite waard was en hoe hij mij op onopvallende wijze aanmoedigde er aan te blijven werken. Verder wil ik Ragnhild Davidse bedanken voor haar hulp bij het uitvoeren van de meta-analyse. Divera Twisk heeft mij op diverse manieren geholpen over mijn 'writer's block' heen te komen tijdens de zo belangrijke eindfase van het schrijven aan dit proefschrift; hiervoor ben ik haar bijzonder dankbaar. Tenslotte dank ik Coen Ederveen voor zijn hulp bij het redigeren van de tekst en Tessa Fagel voor haar idee voor het omslag-ontwerp.

Curriculum vitae

Marjan Hagenzieker werd op 4 maart 1962 in Den Haag geboren. Na het behalen van het Gymnasium β diploma aan de Scholengemeenschap 't Loo in Voorburg, studeerde zij psychologie aan de Rijksuniversiteit Leiden. In 1983 behaalde zij het kandidaatsexamen psychologie. In 1987 werd het doctoraalexamen behaald (cum laude) met als hoofdvak Psychologische Functieleer en als bijvakken Methoden en Technieken van Sociaal Wetenschappelijk Onderzoek en Verkeerskunde (aan de Technische Universiteit Delft). Sinds 1987 werkt zij bij de Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV. Zij heeft onder meer onderzoek gedaan op het gebied van visuele waarneming en aandacht in het verkeer, het gedrag en de verkeersveiligheid van fietsers en bromfietsers, en het beïnvloeden van verkeersgedrag door middel van beloningen. Het onderzoek over dit laatste onderwerp mondde uit in dit proefschrift.