Applied Cognitive Psychology, *Appl. Cognit. Psychol.* **25**: S4–S15 (2011)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/acp.1778

APPLIED COGNITIVE PSYCHOLOGY

25TH ANNIVERSARY

From the archive: 'Time went by so slowly: Overestimation of event duration by males and females' by E. F. Loftus, J. W. Schooler, S. M. Boone, & D. Kline (1987). *Applied Cognitive Psychology*, *1*, 3–13 with commentary

Nominated by Robyn Fivush – Associate Editor

This paper appeared in the very first volume of the very first issue of *Applied Cognitive Psychology* and exemplifies the place of this journal in the field. The authors present an elegant theoretical framework about the study of time from philosophical and psychological traditions, and discuss how time and memory are interwoven. In a series of experiments, Loftus et al. examined participants' perception of time as a function of arousal level, and showed that arousal increases time estimations, especially for females. Throughout the manuscript, the authors knit together theoretical ideas with forensic implications, reinforcing the mission of *Applied Cognitive Psychology* to present the best science that also has profound implications for real world issues and applications.

APPLIED COGNITIVE PSYCHOLOGY, VOL. 1, 3–13 (1987)

Time Went by so Slowly: Overestimation of Event Duration by Males and Females

ELIZABETH F. LOFTUS, JONATHAN W. SCHOOLER, STANLEY M. BOONE and DONALD KLINE

Department of Psychology, University of Washington, Seattle, Washington 98195, U.S.A.

SUMMARY

In three experiments, 469 subjects watched a short videotape of a bank robbery and later estimated the duration of the tape. Subjects invariably overestimated the durations. Accuracy of time estimation was unrelated to amount of free recall (Experiment 1) or accuracy of memory (Experiment 2). Females overestimated to a greater degree than males (Experiments 2 and 3). A more stressful version of the event produced greater overestimates than a less stressful version (Experiment 3). The relationship between induced arousal and time estimation appears to be different for men and women.

The experience of time has been of continuing concern to scholars since the period of the ancient Greeks. This is probably so because time is one of the compelling and universal experiences of our lives (Ornstein, 1969). In recent years the literature on time estimation has grown, with much of the work focusing on estimation as a function of either personality characteristics (such as anxiety) or experimental conditions (Fraisse, 1963, 1984; Meade, 1966; Sarason and Stoops, 1978).

One important aspect of the experience of time, from a theoretical as well as practical perspective, occurs when people judge the durations of events after those events have terminated. This research is about such retrospective estimation, as opposed to prospective estimation. Retrospective estimates of duration have been shown to depend upon many variables, including the complexity and interest value of the material that filled the duration (Fraisse, 1984; Ornstein, 1969; Schiffman and Bobko, 1974). One hypothesis proposed to account for variation in duration estimation following different activities is the 'storage size hypothesis', according to which memory of an event's duration depends upon the amount of information stored in memory about that event. The more information that is stored, the greater is the storage space utilized, and the longer the interval seems to have been (Ornstein, 1969). The storage size hypothesis has received a great deal of support, although an occasional finding, such as that excessive complexity sometimes leads to shorter estimates than more moderate complexity, runs counter to it (Hogan, 1978).

One situation in which accurate duration estimation can be crucial is that in which a witness perceives a complex event and then testifies about that event in court. Time estimates were crucial, for example, in a 1974 case involving a young woman who had killed her boy friend (Loftus, 1974). The prosecutor charged first-degree murder; the defence claimed self-defence. The trial arose out of an incident in which a heated

0888–4080/87/010003–11$05.50

Received 23 December 1985
Revised 21 April 1986

argument began between the woman and her boy friend; she ran to the bedroom, grabbed a gun, and shot him six times. At the trial a dispute arose about the time that had elapsed between the grabbing of the gun and the first shot. The defendant and her sister said 2 seconds, while another witness said 5 minutes. The exact amount of elapsed time made all the difference in the world to the defence, which insisted the killing had occurred suddenly, in fear, and without a moment's hesitation.

Duration estimates are routinely given in cases in which jurors must judge the reliability of an eyewitness account. It is a matter of common sense, as well as a psychological finding (e.g. Laughery, Alexander and Lane, 1971) that the longer a witness has to look at something, the more accurate the memory is likely to be. Jurors who know this may take seriously the estimate of time given by the witness, and use that estimate to judge the likely reliability of the recollection. Thus it is desirable that such estimates be accurate; or, if they are systematically not, then some adjustment for that systematic bias must be taken into account.

When it comes to complex events, do people overestimate time durations as they do in experiments involving relatively pallid materials? In fact, such overestimation has been routinely shown (Buckhout, 1977; Johnson and Scott, 1976; Marshall, 1966). For example, Marshall (1966) asked subjects to watch a 42 s film in which a young man rocks a baby carriage and then flees when a woman approaches him. A week after the subjects had seen the film they were asked how long it had lasted. On the average the subjects thought it had lasted about $1\frac{1}{2}$ min. In another study, subjects saw a staged assault on a university campus. The entire event was recorded on videotape, and lasted 34 s. Some time later subjects were asked questions, including one about the duration of the incident. The average estimate was 81 s, indicating a degree of overestimation approaching a factor of almost $2\frac{1}{2}$ to 1 (Buckhout, 1977). In these studies, time estimation has not been the main focus of the work, but rather something that was examined as a more casual, secondary interest. Still the observation of overestimation is a curious one, especially in light of the fact that some subjects will complain that the simulated incidents went too fast for them to be expected to remember what happened.

Can the storage size hypothesis account for these overestimations. In this article we present new duration estimation data that cannot be readily accounted for by the storage size hypothesis. In Experiment 1 we showed subjects a short videotape of a bank robbery. We observed substantial overestimation of duration of the tape, and the degree of overestimation was unrelated to amount of detail recalled. A storage-size hypothesis might predict that subjects who recalled more details from the event would have produced longer time estimates.

In Experiment 2 we again observed substantial overestimation of duration, and the degree of overestimation was unrelated to accuracy of memory as assessed by a questionnaire. In this study we also found a strong gender difference in duration estimation: females overestimated duration to a greater degree than did males, even though they did not store more accurate information in memory about the event.

To shed further light on the gender difference we conducted Experiment 3, an experiment designed to determine if level of stress or arousal played a role in mediating the gender differences in the duration estimation task. Past research on duration estimation has shown that anxiety is an important variable: high-test-anxious subjects' time estimates have been shown to be significantly greater than the estimates of low-test-anxious subjects (Sarason and Stoops, 1978). This result is thought to be analogous to the tendency to exaggerate the time spent in the dentist's waiting room: anticipating

and experiencing unpleasant, frightening, or threatening events seems to have taken up a lot of time. This observation leads naturally to a hypothesis to account for the gender difference in time estimation: relative to males, females were simply made more anxious by the film, and the higher degree of anxiety caused the film to seem longer. As we shall see, the picture is actually more complicated than this.

EXPERIMENT 1

Method

Subjects

A total of 66 subjects participated in this study. All were students at the University of Washington who participated for course credit. They were run in small groups.

Design and procedure

The subjects came into a laboratory room and watched a simulated bank robbery on videotape. The tape was explicitly designed to be used in research on eyewitness testimony, and was produced at the University of British Columbia, under the direction of Dr John Yuille. The version shown in this study lasted 30 s. After watching the tape, the subjects engaged in some filler tasks lasting approximately 10 min. They were excused and asked to return 48 h later for some allegedly unrelated tasks. The purpose of the filler tasks on the first day was to minimize the likelihood that subjects would think the slide sequence was the single central aspect of the study and to minimize the likelihood that they would rehearse it over the 48 h retention interval.

When the subjects returned to the laboratory they were asked a few routine questions (year in school, future occupation). Information on subject gender was not collected.

After the preliminary questions, subjects were asked to estimate the duration of the tape. Finally they were asked to recall the event in the tape in as much detail as possible. The rationale for asking the time estimation question prior to free recall, rather than the other way around, was to minimize the chances that subjects who first recalled a large number of details would then conclude that the event must have lasted longer.

Results

Many subjects responded to the time question with a 'ballpark' estimate (e.g., 60 s, 1 min, 120 s, 2.5 min). All estimates were converted to seconds for further analysis. The average estimate of duration across 66 subjects was 147.3 s, which confirms earlier findings of time overestimation for complex events. Only two (3 per cent) subjects estimated a duration that was equal to or less than the true value of 30 s.

In order to assess the relationship between time estimation and free recall, each subject's written account was converted to a number that represented the number of details recalled in the account. Subjects were given one point for each object they mentioned that was in fact present in the tape. They were also given one point for each major action that they correctly reported. They were not penalized for incorrectly recording minor details, such as a piece of clothing on the bank robber. The researcher who converted the free recall account was unaware of the time estimate that the subject

had produced. The number of details recalled was then correlated with time estimate across the 66 subjects. That correlation was 0.05, and not significant.

To properly interpret the absence of a significant correlation between two variables, we must show that the two variables are sufficiently reliable to detect a relationship if it were to exist. Thus we asked another researcher to independently convert the free recall protocols to a single number reflecting the amount of free recall. The correlation between the number of details recalled (obtained by the second researcher) and time estimate was −0.02 across 66 subjects. The agreement between the two independent researchers' estimates of amount recalled was 0.87. Thus it is reasonable to conclude that the lack of relationship between amount recalled and time estimate is not due to a lack of reliability in scoring details recorded.

Discussion

Contrary to a plausible prediction from a storage-size hypothesis, time estimation was unrelated to amount of free recall. It is still possible that the free recall measure was not the best one to equate with storage size. Perhaps subjects were inhibited from demonstrating all that they knew because of a reluctance to expend much energy writing details down on paper. A different measure of the strength of memory which required less effort to demonstrate could conceivably correlate with time estimation. This idea was explored in Experiment 2.

EXPERIMENT 2

This study attempted to replicate the time overestimation phenomenon observed in Experiment 1. In this study we also examined the relationship between time estimation and accuracy of memory for details. It is conceivable some subjects might be distracted from paying full attention to the tape. These individuals would be expected to perform relatively poorly on a memory test, and they might also produce lower time estimates. If this occurred there would be a positive correlation between time estimation and test performance. Such a result would be consistent with a storage-size hypothesis.

Method

Subjects

A total of 266 subjects—110 males and 156 females—participated in this study. All were students at the University of Washington, who participated for course credit. They were run in small groups.

Design and procedure

The subjects came into a laboratory room and watched a 30 s simulated bank robbery on videotape. After watching the tape, the subjects engaged in some filler tasks lasting approximately 10 min. They were excused and asked to return 48 h later.

When the subjects returned to the laboratory they received a 25-item questionnaire concerning the videotape. One of the questions asked for an estimate of the duration of the videotape. The remaining questions asked for recall of the details of the tape.

Results

The average estimate of duration across 266 subjects was 152 s, which replicates Experiment 1 closely. The correlation between time estimation and accuracy on the remaining test items was −0.02 and not significant.

When the data were analysed separately for males and females, we found that males estimated an average of 133.14 s (range = 15 to 480 s), while the females estimated an average of 165.0 s (range = 10–900 s), $t(264) = 2.23$, $p < 0.025$. Even when the data from five females (3.2 per cent) who produced inordinately long estimates ($\cong$900 s) were removed from the analysis, a significant sex difference remained.

The distribution of responses for males and females is shown in Figure 1. As can be seen, only 6.4 per cent of males and 0.6 per cent of females estimated a duration that was equal to or less than the true value of 30 s. The vast majority of subjects produced overestimates.

Longer time estimates tended to be reported in minutes rather than seconds (e.g. 5 min, 10 min), and these were converted to seconds for purposes of analysis. Males and females did not differ in the degree to which they reported their time estimates in seconds (e.g. 60 s) or in minutes (e.g. 1 min) when comparable times were estimated.

The storage size hypothesis might account for the observed overestimation by proposing that females stored more information in memory about the event, and this is why the interval seemed to have been longer. A finding that females performed better on the remaining 24 test items would provide evidence for this hypothesis. However, no differences were observed as a function of gender on the remaining items of the test. Females correctly answered 15.5 questions correctly compared to 15.2 for males, $p > 0.4$. As a point of interest, the five females who produced extraordinarily long time estimates ($\cong$ 900 s) answered an average of 15.8 questions correctly.

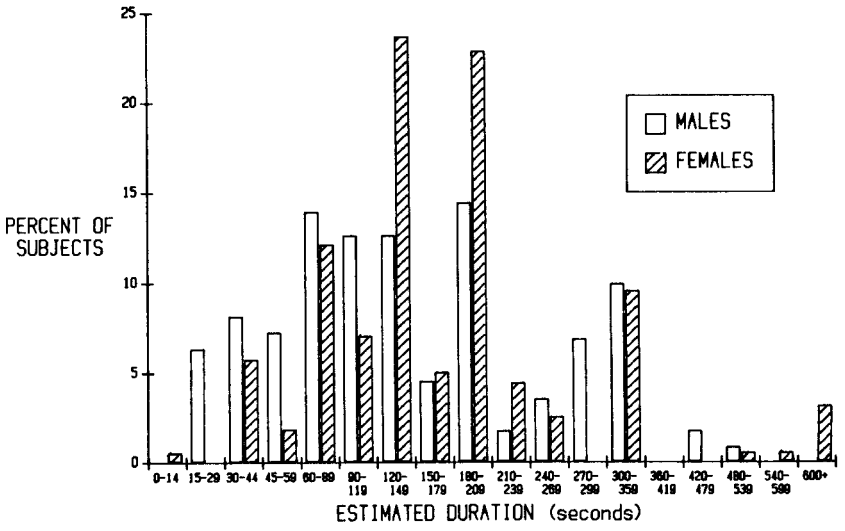

Figure 1. Distribution of duration estimates in Experiment 2 for males and females

Discussion

Subjects invariably overestimated the duration of the bank robbery tape. The finding that females overestimated the duration to a greater degree than males was initially surprising, given that many studies of time estimation using more pallid materials have not observed gender differences (Schiffman and Bobko, 1974; Sarason and Stoops, 1978). Why do females overestimate to a greater degree in this case? One hypothesis is that the effect is associated with differential degrees of stress or arousal experienced as a result of watching the tape. To test this notion, Experiment 3 was conducted.

EXPERIMENT 3

Method

Subjects

A total of 137 subjects participated in this study—68 males and 69 females. All were students at the University of Washington who participated for course credit. They were run in small groups.

Design and procedure

The subjects watched one of two versions of a videotaped bank robbery lasting 28 s. These versions came from the same source of tapes as did the version used in Experiment 1. A 'low-stress' version depicted the robber entering the bank and calmly handing a note to the teller. A 'high-stress' version depicted the robber displaying an automatic pistol, and using profane and threatening language. Approximately half of the males and half of the females saw each version of the tape.

After the videotape was shown, subjects filled out a 12-item questionnaire which included a question asking for an estimate of the duration of the tape. In addition, all subjects were asked how upsetting they found the videotape to be, using a scale from 1 = not upsetting to 6 = highly upsetting. To ensure that subjects would rate their own specific reaction to the tape, rather than judge the tape's content as upsetting in the grand scheme of things, we deliberately asked 'How upsetting did you find the videotape to be?'

Results

In all conditions, subjects overestimated the duration of the event. Mean time estimates of males and females who viewed the low- and high-stress versions are shown in Table 1. The mean estimates are lower than in the previous experiments, undoubtedly because testing occurred soon after the videotape in this study rather than after a 48 h retention interval. A two-way analysis of variance indicated that females produced longer estimates than males, $F(1,133) = 9.91$, $p < 0.005$. Moreover, the high-stress version produced higher estimates than the low-stress version, $F(1,133) = 4.36$, $p < 0.05$. The interaction was not significant, $F < 1$.

Mean values indicating degree of reported arousal or distress due to the tape are shown in Table 2. Females reported higher degrees of arousal than males, $F(1,133) = 6.09$, $p < 0.02$. The high-stress version produced higher reported arousal than did the

Table 1. Mean time estimation (seconds) in Experiment 3

	Gender	
	Male	Female
Induced stress		
Low	50.6	73.2
High	64.2	95.1

Table 2. Mean arousal reported in Experiment 3

	Gender	
	Male	Female
Induced stress	1.75	1.94
	2.06	2.76

low-stress version, $F(1,133) = 1.99$, $p < 0.01$. The interaction was not significant, $F(1,133) = 1.99$, $p > 0.10$.

To assess the relation between reported arousal and duration estimate we calculated the correlation between these two variables. The overall correlation was −0.04, and was not statistically significant. We further examined the correlations between the two dependent measures separately for each condition. These data are shown in Table 3. None of the correlations is significantly different from zero.

Time estimates were examined for males and females who reported the same degree of arousal. Only at reported arousal levels of 1 and 2 were there sufficient numbers of males and females to make meaningful comparisons. At both arousal level 1 and 2, females produced higher time estimates than males. At reported arousal level 2, males

Table 3. Data from Experiment 3

Condition	*n*	*r*	Mean time	Mean arousal
Males				
Low stress	32	−0.23	50.63	1.75
High stress	36	−0.16	64.17	2.06
All males	68	−0.15	57.19	1.91
Females				
Low stress	35	−0.22	73.23	1.94
High stress	34	−0.10	95.35	2.76
All females	69	−0.07	84.13	2.35
Males + Females				
Low stress	67	−0.18	62.43	1.85
High stress	70	−0.03	79.31	2.40
All subjects	137	−0.04	71.05	2.13

produced a mean time estimate of 48.2 while females produced a mean time estimate of 93.4, $t(51) = 3.24$, $p < 0.01$. At arousal level 1, males produced a mean time estimate of 69.0 while females produced a mean time estimate of 85.6, a result that is in the same direction but falls short of statistical significance, $0.10 > p > 0.05$.

Performance on the remaining 11 items of the questionnaire did not vary as a function of gender. Females produced a mean score of 4.6 items, while males produced a mean of 4.9 items, $p > 0.5$. Even when gender was examined in the low-stress and high-stress conditions separately, performance was equivalent, $p > 0.5$.

GENERAL DISCUSSION

Taken together, the three experiments show pervasive overestimation of the duration of the videotape. While both sexes overestimated the duration, females overestimated to a greater degree than males. Moreover, Experiment 3 revealed that subjects who watched a more stressful version of the event produced longer time estimates than those who watched a less stressful version.

To account for these results we have relied upon a method of testing simple theories of causation articulated by Bamber (1979). We begin by expressing a functional relationship between an independent variable (stressfulness of the film) and two dependent variables (reported arousal and time estimate). The data from Experiment 3 tell us that changes in the independent variable produced changes in both dependent variables. In Bamber's (1979) terms, this is an observational structure, as is shown in Figure 2.

We suggest the existence of a latent (hypothetical unobservable) variable, which we term internal level of arousal. The value of the latent variable depends in part on the value of the independent variable; that is, internal arousal will depend upon film-induced stress level. For each person, internal arousal is a function of induced stress. The dependent variables, reported arousal and the estimated time duration, are functions of internal arousal. This basic theoretical structure is illustrated in Figure 2.

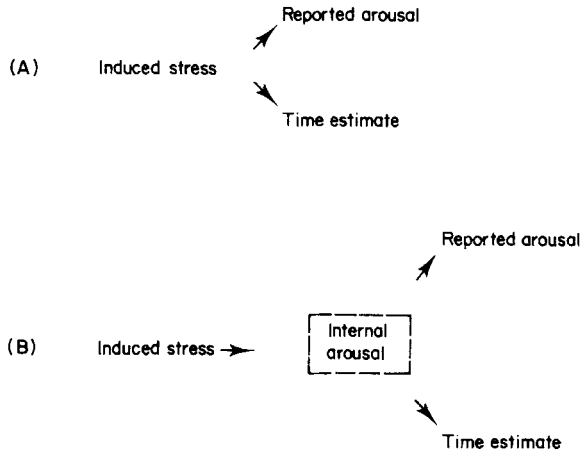

Figure 2. An observational structure and corresponding theoretical structure for describing the relationship between induced stress and both reported arousal and time estimation.

The simple model rejected by the present results is that males and females have the same function relating induced stress to internal arousal, and internal arousal to the dependent variables, reported arousal and time estimation. Our results suggest that the same degree of film-induced stress produces more internal arousal for females. Moreover, it may also be the case that internal arousal is differentially related to reported arousal. Both hypotheses are plausible. The same degree of induced stress could produce more arousal for females because females have far less experience with and preference for watching violent television than do males (Parke and Slaby, 1983). These sex differences have been reported for a wide range of age levels, from preschool through high school. The same level of internal arousal could lead to differences in reported arousal if, for example, males were more reluctant to admit to high degrees of arousal than females.

Some readers may be puzzled about how experimentally inducing stress could produce changes in time estimation even though no overall correlation between reported arousal and time estimation was observed. An analogy to the pulse rate of runners on a treadmill may help to clarify this issue. If we take a group of people of varying levels of fitness and ask them to jog for 20 min, we may not observe a correlation between pulse rate and miles run. While for any given level of fitness we would expect a higher pulse rate for further distances, across levels of fitness the pattern would be reversed. Unfit people would be expected to run less far and have higher pulse rate, while fit people would be expected to run further with a generally lower pulse rate. Thus, across subjects we would not necessarily expect to see positive correlations between pulse rate and distance run on a treadmill.

If, on the other hand, we were to experimentally control the duration that two randomly selected groups were allowed to run, we would expect to observe a relationship between pulse rate and distance. Specifically, if we divided our group in two, and had half of our subjects run for 10 min and the other half run for 20 min, we would expect the latter group to exhibit a higher pulse rate. This prediction follows because, by experimentally controlling amount of exertion, we no longer confound exertion with level of fitness, as would be the case in a purely correlational comparison.

The treadmill analogy nicely demonstrates how it is possible to observe a relationship between two variables when one of the variables is controlled experimentally, but not when two variables are simply correlated within a general population. Applying the treadmill analogy to the relationship between arousal and time estimation suggests that there must exist some other individual difference variable, analogous to physical fitness, which confounds the experimentally observed relationship between arousal and time estimation. While it is clearly beyond the scope of this study to identify such a variable, it is easy to imagine one. For example, it could be that people who tend to be easily aroused respond to the same internal state of arousal with lower reported arousal than subjects who are not easily aroused. Intuitively such a prediction makes sense; for people who are easily aroused, a highly aroused state will not seem unusual and consequently should not produce relatively high arousal reports. Accordingly, the correlation between arousal and time estimation could be confounded if easily aroused subjects tended to respond to the same internal state with lower arousal estimates than uneasily aroused subjects. Clearly, more research is necessary in order to develop a more complex theoretical understanding of the various factors that contribute to the complex relationship between internal arousal, self-reported arousal, and time estimation.

While still requiring further research, the current results have suggestive practical

significance. Given two persons—John who reports that an event was highly upsetting and Jim who claims it was not—we cannot assume that John is likely to produce a longer estimate of the event than Jim. However, it is reasonable to presume that both John and Jim would produce higher estimates for an objectively stressful event than the estimates they respectively produce for a non-stressful event of the same duration. Moreover, they are likely, on the average, to produce lower estimates than their female counterparts.

An elucidation of the basic processes involved in time estimation in general, and retrospective overestimation more specifically, can help us to understand errors in estimation made by actual witnesses and victims to real events. In many instances when a witness gives an estimate regarding the duration of some event, it is not possible to know the actual figure; thus one cannot be certain about the degree of overestimation. There are, however, some 'real-world' exceptions. In one study (Schneider, Griffith, Sumi and Burcart, 1978), 212 reports of crime incidents from a Portland, Oregon, victimization survey were matched with official crime reports of the same incidents. Among other facts, respondents in the survey were asked to estimate the amount of time it took before the police arrived on the scene. There were only two survey respondents who estimated the time to be shorter than what police records showed. Almost half the respondents estimated the time within 15 minutes of the estimate given on the police report, and the other half of the respondents said that the time was at least 15 min longer than indicated by the police report. It is of interest to note that about 10 per cent of respondents gave estimates that were dramatically longer (over 2 h) than indicated by the police report. While the possibility exists that the police have underestimated the time, in Portland this possibility is very remote. The victim's call to the police is recorded, the dispatcher's call to the officer is recorded, and the officer's call that he has arrived on the scene is recorded. Estimates are made in seconds, not just in minutes. In short, it is highly likely that the victims are overestimating, and sometimes quite dramatically.

Given that crime victimizations are undoubtedly quite stressful events for most people, the routine overestimation should not be surprising. It would have been interesting to know whether the same gender differences observed in laboratory estimates of time were also apparent in the real-life estimation obtained in Portland, but such information is not reported.

ACKNOWLEDGEMENT

This research was supported by grants from the National Science Foundation and from the National Institute of Mental Health. We are grateful to G. Loftus and J. Palmer for helpful discussions about theoretical causation structures. Please address all correspondence to E. Loftus, Department of Psychology, University of Washington, Seattle, Washington, 98195.

REFERENCES

Bamber, D. (1979). State–trace analysis: a method of testing simple theories of causation. *Journal of Mathematical Psychology*, **19**, 137–181.

Buckhout, R. (1977). Eyewitness identification and psychology in the courtroom. *Criminal Defense*, **4**, 5–10.
Fraisse, P. (1963). *Psychology of time*. New York: Harper & Row.
Fraisse, P. (1984). Perception and estimation of time. In *Annual Review of Psychology*. Palo Alto: Annual Reviews, pp. 1–36.
Hogan, H. W. (1978). A theoretical reconciliation of competing views of time perception. *American Journal of Psychology*, **91**, 417–428.
Johnson, C. and Scott, B. (1976). 'Eyewitness testimony and suspect identification as a function of arousal, sex of witness and scheduling of interrogation'. Paper presented at the American Psychological Association, Washington D.C.
Laughery, K. R., Alexander, J. E. and Lane, A. B. (1971). Recognition of human faces: effects of target exposure time, target position, post position, and type of photograph. *Journal of Applied Psychology*, **55**, 477–483.
Loftus, E. F. (1974). The incredible eyewitness. *Psychology Today*. December.
Marshall, J. (1966). *Law and psychology in conflict*. New York: Bobbs-Merrill.
Meade, R. D. (1966). Achievement motivation, achievement, and psychological time. *Journal of Personality and Social Psychology*, **4**, 577–580.
Ornstein, R. E. (1969). *On the perception of time*. New York: Penguin Books.
Parke, R. D. and Slaby, R. G. (1983). The development of aggression. In P. H. Mussen (Ed.), *Handbook of child psychology*, (4th edition, volume 4). New York: Wiley.
Sarason, I. G. and Stoops, R. (1978). Test anxiety and the passage of time. *Journal of Consulting and Clinical Psychology*, **46**, 102–109.
Schiffman, H. R. and Bobko, D. J. (1974). Effects of stimulus complexity on the perception of brief temporal intervals. *Journal of Experimental Psychology*, **103**, 156–159.
Schneider, A. L., Griffith, W. R., Sumi, D. H. and Burcart, J. M. (1978). *Portland forward records check of crime victims*. Washington D.C.: U.S. Dept. of Justice.

