

P.J. van Koppen & S.K. Lochun (1997) Portraying perpetrators: The validity of offender descriptions by witnesses. *Law and Human Behavior*, 21, 663-687

## Portraying Perpetrators: The Validity of Offender Descriptions by Witnesses

Peter J. van Koppen<sup>1,3</sup> and Shara K. Lochun<sup>2</sup>

---

*Using archival data from official police records, we assessed how much information witnesses include in their descriptions of the offenders, how accurate this information is, and which factors influence the completeness and accuracy of these descriptions. The study included 2,299 descriptions by 1,313 actual forensic witnesses given of 582 different robbers involved in commercial robberies in The Netherlands. Witnesses tend to provide little information in their descriptions. They describe more general features, such as sex, race, and build, than facial characteristics. The descriptions were compared to the real appearance of the robbers. Although the witnesses gave little information, the little information they gave tended to be relatively correct. The accuracy of the descriptions could hardly be predicted from completeness of the descriptions, nor from witness or robbery characteristics.*

---

### INTRODUCTION

The wiser part of the robber community takes precautions against recognition. They use all kinds of garments and accessories to mask their features, even if they rob commercial targets which do not have a security camera. Some turn to balaclavas or stockings or both, others use a fancy dress mask, and still others stick to a simple but effective baseball cap and sunglasses. They know that the police routinely not only ask witnesses to describe what happened, but also to give a description of the perpetrator and, after a suspect has been apprehended, to attempt an identification. Thus, eyewitnesses play a central role in the criminal justice system, both in the apprehension and the conviction of offenders.

For centuries, and particularly since the turn of the century, both the description of what happened (Loftus, 1979) and identification by witnesses have been

<sup>1</sup>Netherlands Institute for the Study of Criminality and Law Enforcement (NISCALE), Leiden, The Netherlands.

<sup>2</sup>Department of Psychology, University of Exeter, United Kingdom; present address: School of Health and Social Welfare, The Open University, Walton Hall, Milton Keynes, MK7 6AA, United Kingdom.

<sup>3</sup>To whom correspondence should be addressed at Netherlands Institute for the Study of Criminality and Law Enforcement (NISCALE), Wassenaarseweg 72, 2333 AL Leiden, The Netherlands (e-mail: VanKoppen@NISCALE.leidenuniv.nl).

topics of psychological research (Cutler & Penrod, 1995; Cutler, Penrod, & Martens, 1987; O'Rourke, Penrod, Cutler, & Stuve, 1989; Shapiro & Penrod, 1986; Sporer, 1982; Wells, 1993). Although offender descriptions by witnesses are an important source of information in the investigation of crimes (Cutshall & Yuille, 1989; Geiselman, Fisher, MacKinnon, & Holland, 1985; Köhnken, Schimossek, Aschermann, & Hofer, 1995), little psychological research has been directed to such descriptions (Farrington & Lambert, 1993; Sporer, 1996; Tollestrup, Turtle, & Yuille, 1994). Thus, little is known of their validity. In the present study the completeness and validity of offender descriptions are assessed, using archival data. Specifically, we address the following three questions: (1) How much information do witnesses include in their descriptions, and how accurate is this information? (2) What factors influence the completeness and accuracy of these descriptions? and (3) Can description accuracy be predicted from description completeness?

### Using Archival Data

In experimental studies of witness statements it is possible to include both errors of commission and omission in experimental assessments of accuracy. Laboratory studies in particular are difficult to generalize to real-world situations, since the validity of the materials, subject populations (most often undergraduate students), and tasks used can all be questioned (Konecni & Ebbesen, 1986; McCloskey & Egeth, 1983; McKenna, Treadway, & McCloskey, 1992). Simulated crimes and field studies provide a more realistic representation of an actual eyewitness situation and allow offender descriptions to be compared against a known 'reality' (Wells, 1993; Yuille, 1993; see also Davies, 1992, 1995; Davies, Tarrant, & Flin, 1989; Macleod & Shepherd, 1986; Tollestrup et al., 1994). Thus, archival studies form a useful supplement to experimental studies in this field.

The validity of assessing the accuracy of archival material, however, is also debatable. Sporer (1996), for example, claims that verifying the contents of witness descriptions in archival analyses is generally impossible, since we rarely know for sure who the perpetrator of any crime was (but see Cutshall & Yuille, 1989, for exceptions). We suggest that accuracy can be estimated for crimes for which a suspect has been convicted only if two assumptions are accepted: (1) The persons described by the witness are the persons convicted of the offence, and (2) The suspects have not significantly changed their appearances (apart from clothing) between when the crime was committed and when they were apprehended and their appearances recorded in police files.

There are problems associated with both of these assumptions. For instance, the true incidence of false convictions is very difficult to assess (see Huff, Rattner, & Sagarin, 1996), and the appearance of an offender may change considerably over a short period of time for a number of reasons, such as drug abuse, weight gain or loss, the addition or removal of facial hair, and the use of hair dye. The prevalence of these sources of error is virtually impossible to quantify and predict. We therefore do not claim that this methodology will produce data that are in some way 'better' than those obtained using more controlled paradigms (cf. Yuille, 1993),

but it is increasingly being acknowledged that reliable findings will only emerge if common results are identified across a variety of different methodologies (Davies, 1992; Yuille, 1993; Yuille & Wells, 1991). Thus, the present archival study of descriptions of offenders can be an addition to the field. We return to the assumptions underlying archival research in the discussion section.

### Completeness and Accuracy of Descriptions

Witnesses can only give a good description if they had the opportunity to observe the offender clearly. In a forensic context ability to view the perpetrator may be considerably poorer for many witnesses than can be simulated in a laboratory study. The offender may be moving, may have his or her back to the witnesses, or may be some distance from them. Thus witnesses may be able to observe only the 'coarser' features of the offender (height, sex, etc.), rather than facial features, and among facial features witnesses may remember coarser aspects (hair, skin, and face shape) rather than the 'finer' details. These factors suggest that predictions of description completeness and accuracy taken from laboratory studies should be applied with some reservation to forensic witness performance.

Experimental research has, however, provided indications of the type of information we might expect to obtain from forensic witnesses. Research on the relative saliency of different facial features (Ellis, Deregowski, & Shepherd, 1975; Ellis, Shepherd, & Davies, 1980; Shepherd, Ellis, & Davies, 1977) has shown a general preference for upper as opposed to lower facial features in descriptions, with hair and eye detail mentioned significantly more often than nose, mouth or chin. The order of frequency with which different facial characteristics appeared in a large number of descriptions was remarkably consistent across studies, being (highest first): hair, eyes, nose, eyebrows, face shape, chin, lips, mouth, ears, face lines, complexion, forehead, and cheeks.

Sporer (1992) found a slightly different pattern in an archival study. Face shape and skin descriptors were mentioned more than twice as often as any other facial descriptor (with the exception of beard and glasses). Sporer's analysis also illustrated that with the exception of the offender's hair, general descriptors such as age, height, and build were mentioned more often than any facial attribute, and race was mentioned more frequently than all facial features apart from beard and glasses, while the most commonly reported characteristic was the clothing of the perpetrator (see also Farrington & Lambert, 1993; Kuehn, 1974).

Only a few studies compared witness descriptions to official police reports of how the offender really looked (Farrington & Lambert, 1993; Tollestrup et al., 1994). Witness descriptions of sex, ethnicity, and distinguishing facial characteristics were almost always accurate. That is not completely surprising, since these can be considered to be very salient aspects of a person's physical appearance, and it has been demonstrated that, at least for sex, people need very little information in order to make accurate judgments (Barclay, Cutting, & Kozłowski, 1978). Witnesses of robberies may also use default values for the sex of the perpetrator (Lindsay, Martin, & Webber, 1994), since almost all robbers are male, at least in the Netherlands.

The descriptions of hair color, hair length, and accent tend to show lower agreement with recorded appearance. Accuracy for age, height, and weight were even poorer, mainly because witnesses overestimate these features (for similar findings from experiments see Flin & Shepherd, 1986).

Descriptions are of course never complete. They can only be measured in terms of the number of distinct features mentioned by witnesses. In prior research the mean number of features mentioned varied around 10 (Sporer, 1992; Tollestrup et al., 1994), and witnesses mentioned fewer clothing descriptors than appearance descriptors. The latter distinction is useful from the perspective of the police. In the first few minutes after a crime has taken place a description may be used to try to apprehend a perpetrator while he or she is still in the immediate vicinity. At this point, the most salient aspect of offenders is likely to be their clothing. However, clothing can easily be discarded and destroyed within minutes of a crime taking place, and thus these details become less relevant for apprehending the criminal if the description is given more than a couple of minutes after the robbery has occurred (Farrington & Lambert, 1993; Sporer, 1992). Such details, though, may still be important to add to the circumstantial evidence in court against a defendant if the items of clothing described in detail by witnesses are found among the possessions of a suspect. Given their varying importance, we shall evaluate these two types of descriptive information independently.

The ability to predict description accuracy or usefulness from measures of completeness alone, particularly from the length of a description, would obviously be a great benefit to police officers in weighing information contained in different descriptions. Previous research suggests, however, that such a relationship is unlikely to exist (Wells, 1985; Yuille & Cutshall, 1986). Farrington and Lambert (1993) point out that as the number of characteristics included in a description increases, so does the probability of one being incorrect, intuitively suggesting a negative relationship between completeness and accuracy.

#### Factors Affecting Eyewitness Descriptions

Perhaps the most extensively studied predictor of witness performance has been the threat experienced by the witness. In a review of 21 experimental studies of the effect of arousal on memory, Deffenbacher (1983) reported that 10 of those studies demonstrate that higher arousal levels increase witness accuracy, or at least do not decrease it, while a further 11 show the converse effect: lower accuracy with increased arousal. The debate seems to end at a draw, but the divergent results may also be explained by the variations in definition and indicators of threat (compare, e.g., Christianson & Hübner, 1993; Macleod & Shepherd, 1986). In archival studies the presence of a weapon seems to enhance, rather than reduce the number of details in witness descriptions (Sporer, 1992; Tollestrup et al., 1994).

It can be expected that, apart from the influence of arousal, descriptions will be less complete and less accurate if there is more to describe, and better if conditions for viewing the offenders are better. So it can be expected that robberies with more robbers decrease accuracy and completeness (Clifford & Hollin, 1981),

but Sporer (1992) failed to find this effect. He did find that good lighting produced more complete descriptions (compare Wagenaar & Van der Schrier, 1996; Yarmey, 1986). It is intuitively viable to expect that the longer something can be observed, the more complete and accurate recall of it will be. Thus, it is of interest to consider the influence of witnessing duration on the completeness and accuracy of descriptions. Witness estimates of crime duration have been suggested to reflect the length of observation of the perpetrator (Pigott, Brigham, & Bothwell, 1990). Experimental research, however, shows that witnesses tend to overestimate these durations in a nonuniform manner (Alper, Buckhout, Chern, Harwood, & Slomovits, 1976; Loftus, Schooler, Boone, & Klein, 1987; Pigott et al., 1990), suggesting that estimated event duration would not have a consistent effect on description quality. A nonsignificant relationship between description completeness and both duration and opportunity to observe found by Sporer (1992) suggests that this interpretation is likely to be correct.

Sporer also found that disguise did not lead to any significant detriment in completeness. This finding is counterintuitive, when considering completeness of facial descriptions and excluding details of the disguise itself (S. L. Sporer, personal communication, 12 September 1996), and may depend upon the way in which this variable was coded: Although there were many different types of disguises used in Sporer's sample, the overall *N* was such that only a dichotomous distinction was possible (S. L. Sporer, personal communication, 12 September 1996). A dichotomous variable simply recording whether or not an offender wore any form of disguise might be expected to be less sensitive than the continuous measure used in this study, thus it is intuitively predicted that a greater amount of disguise will be associated with less complete descriptions.

Similarly, studies on the influence of witness characteristics on descriptions have produced confusing results. Some studies fail to demonstrate gender differences (Clifford & Scott, 1978; Macleod & Shepherd, 1986), while Sporer (1992) found conflicting trends: males gave significantly longer descriptions, whereas females tended to include more units of information (although this difference was not significant).

In contrast to a vast literature concerning children (Goodman & Schwartz-Kenney, 1992), there is relatively little research on the competency of elderly adults as witnesses. Nevertheless, experimental evidence seems to suggest that older witnesses do not perform significantly differently from other adult age groups (Adams-Price, 1992), but the very elderly perform less well than younger adults (Bornstein, 1995; Yarmey, 1996). In the absence of a reasonable number of child and elderly witnesses, no significant effect of age is expected in the present study.

The extent to which witnesses know a crime is taking place at the time they observe the perpetrator is similarly problematic. Tollestrup et al. (1994) observed that the completeness of descriptions provided by victims of fraud—who at the time of interacting with the perpetrator are usually not aware that they are witness to an important event—was significantly lower than that of the victims and other robbery witnesses who were present during the crime. Although this indicates that knowledge that the person observed is committing a crime might influence description completeness, differences between victims and other robbery witnesses were

neither directionally consistent nor significantly different. Thus the influence of knowledge may be mediated by other factors.

Yuille and Cutshall (1986) suggest that those more involved in a crime are also closer to the perpetrator, indicating a possible confound in using type of witness as a predictor of witness performance. However, the proposed association between involvement and proximity to the perpetrator is not always the case. The present study therefore investigates witness involvement, but considers the influence of possible confounding factors (the distance and type of obstructions between the offender and the witness, and their position relative to each other) independently. In line with the findings of Tollestrup et al., it is predicted that witnesses unaware that they have seen a significant event will provide less complete descriptions than other witnesses, and it is expected by extrapolation that a similar finding will be observed for description accuracy. No previous research could be found relating to the three potentially confounding situational factors; however, intuitively it is expected that greater distance, a poorer relative position, and opaque objects (as compared to transparent objects or no obstruction) in between the witness and the perpetrator will all be associated with less complete and less accurate descriptions.

## METHOD

### Data

The data for this study come from official court records kept by prosecution offices. The case files include all statements recorded by the police, findings of the police investigation, suspect and witness interrogations by the judge-commissioner and by the court at trial, and the written decision of the court. We extracted descriptions given by witnesses from these case files. The data set included almost all robberies for which one or more of the robbers were convicted in The Netherlands by a trial court in 1992. Only the robbers and robberies for which the case files could not be extracted or were missing were not included in the study. The study was aimed at robberies of commercial targets and dwellings, excluding street muggings.

In The Netherlands, almost all criminal trials are conducted using documented evidence, mainly produced by the police; witnesses are rarely questioned in courts. It is a statutory requirement that the police use as much as possible witnesses' own words as given in their statements. This gives the advantage that the police put much energy into preparing the case files. Most witness statements were recorded shortly after the robbery (66% the same day and an additional 14% the next day). The disadvantage, however, is that we could not interview the witnesses ourselves, but had to rely on the work done by the officers who took their statements.

The study included 431 different robberies in which 1,313 witnesses gave 2,299 descriptions of the offenders. Most witnesses saw the robbery itself (82%); the others either witnessed the escape (11%) or the preparations by the robbers (7%).

The robberies most commonly involved two offenders (42% of the robberies), a quarter were committed by a single robber, and the remainder by more than two

offenders, with a maximum of eight robbers in a single robbery. The robberies took place all across the country and included the whole range of possible commercial targets (e.g., banks, post offices, all kinds of shops, restaurants, cafes, and snack bars) and dwellings.

The data on the robbers' true appearances were extracted from the national database on offenders (the so-called Herkenningdienst Systeem) maintained by the police. This database includes written descriptions of the appearances of the offenders as recorded by the police and is updated every time an offender is arrested. Although 582 robbers (aged 14–55 years, and including 15 women) participated in the 431 robberies, we could compare the descriptions of witnesses to the real appearance of only 462 robbers, either because it remained unclear which particular robber was described by a witness, or because their data were not available from the national archive. So, only the 1650 descriptions by witnesses of these 462 robbers could be verified. These offenders ranged in age from 15 to 50 years, and included only 6 women.

#### Independent Variables

Independent variables were comprised of both witness characteristics and robbery characteristics. All independent variable information was copied or estimated directly from the police files unless specifically stated below. Witness characteristics included sex, age, type of witness (preparations, robbery, escape), relation to robbery target (employee, customer, passerby), delay in days between robbery and description, and the number of different descriptions provided by each witness at a robbery. Robbery characteristics included lighting conditions (good or poor), obstructions to the witness's view of the suspect (none, glass, opaque objects), duration of perception (short, medium, long), estimated distance from the robber (in centimeters), position of the witness relative to the suspect (four ordered levels), estimated level of threat experienced by the witness (16-point scale), and the amount of facial disguise worn by the robber (percentage).

Lighting conditions were estimated from the time of day and whether the witnesses were inside or outside when they saw the robber. Lighting was classified as 'good' for events witnessed outside in daylight or inside well-lit buildings (typically those open only in the daytime), and was categorized as 'poor' for events which took place outside during the evening or night, or inside unlit or dimly lit buildings (e.g., a bar). The length of time for which a witness could have viewed a robber (duration of perception) was estimated from the description of the event described by the witness. Only rarely did the witness statement contain a specific time estimate. Opportunities to observe which were estimated to last 5 minutes or more were classified as long, those which lasted somewhat shorter were classified as medium, and those which lasted 1 min or less were classified as short. The position of the witness relative to a particular robber was estimated from the information contained in the file, resulting in four ordered levels ranging from best to worst in terms of opportunity to view the offender. An example of the 'best' opportunity is



the witness being held by the robber, whereas the 'worst' opportunity included such things as the witness being blindfolded or hiding behind or under furniture.

Of course all robberies involve a certain amount of threat, otherwise the robber would not expect the victims to hand over any money. Still, some robberies are much more violent and threatening to the witnesses than others. We therefore made a threat index, based on information contained in the case file. A witness to the robbery initially scored +3 on this index, a witness to the escape +2, and a witness to the preparations +1. Additionally, 1 was added for each of the following circumstances: people killed, use of a weapon to threaten, use of violence during robbery, people wounded, the taking of hostages, assault, physical restraint of witnesses, shots fired, victims tied up, vandalism, active resistance by robbery victim, victims stunned or intoxicated by robbers, robbers using alcohol or drugs. If there was bulletproof glass between witness and robber, 1 was subtracted.

The degree of disguise worn by each robber had to be estimated from the description given by the witnesses. We defined disguise as anything worn by the offender which obstructed the witnesses' view of their facial features. We computed the degree of disguise from the type of head covering worn by the offender and whether or not the offender was wearing sunglasses. For each of the 38 kinds of head covering (including such things as different types of masks and hats, and combinations such as a baseball cap with a balaclava) the authors consensually determined the percentage of the face that *would not* be covered. If a nylon stocking was used to obscure facial features, the resulting percentage was halved. For sunglasses we subtracted another 20%. In some cases this resulted in a percentage lower than 0. It can be expected, however, that at least some small piece of facial information might be available for view with almost all disguises (for example, a lock of hair, a small patch of skin, the overall shape of the face). Thus, we changed any percentage lower than 5 to an arbitrary 5%.

### Dependent Variables

The descriptions were copied verbatim from the case files. This initial content coding was done by a single person. To be able to compare descriptions to the real appearance of the robbers as coded in the Herkenningsdienst Systeem the verbatim descriptions were divided into the categories used by the police.

The completeness of the descriptions was measured in four ways. We counted the number of text characters in each description. We also counted the number of descriptor units mentioned in each description, out of a possible total of 43 (these are listed in Tables 2 and 3). The two remaining completeness estimates consisted of the number of permanent characteristics that were mentioned in a description and the number of temporary characteristics mentioned, respectively. We defined permanent characteristics as those which cannot be changed in 1 or 2 min by the offender; all others are temporary (see Tables 2 and 3).

Obviously none of the 19 temporary (clothing) characteristics could be compared to the real appearance of the robbers. However, most of the 24 permanent characteristics could be verified against police file information, the exceptions being

**Table 1.** Example of Rules Used for Recoding Witness Descriptions of Age and Height to Accuracy Scores

Witness description (years)	Range given by witness (years)	Range expanded to include target age (years)	Range used for computations (years)	Accuracy score
25	1	25-25	1	3
20-26	7	20-26	7	2.69
20-23	4	18-25	8	2.64
35-38	4	25-48	24	1.82
18-20	3	13-25	13	2.38
Older than 20	Undetermined	20-55 (max)	36	1.21
Younger than 36	Undetermined	12 (min)-36	25	1.77

*Note:* This example refers to the accuracy of the ages produced by witnesses in cases in which the robber in fact is 25 years old.

glasses and earrings. Furthermore, the information coded within two description units in the data file (appearance/race and skin color) corresponded to only a single descriptor in the police file (appearance/race), and therefore the two former variables were combined and compared against the latter in order to determine their accuracy. The total number of verifiable descriptors was therefore 21.

For each of 19 of these matched descriptors a table was generated displaying the categories for that variable recorded in witness statements against the categories contained within the police description files. The two excluded verifiable descriptors were age and height, for which a slightly different procedure was used to ascertain accuracy (see below). Accuracy ratings on a 3-point scale were assigned to each cell in a cross-table. All accuracy ratings were judged by the consensual opinion of the authors. *Wrong* corresponded to a witness description which based on the police file description was completely inaccurate, *partially correct* to a witness description that was partially accurate, and *correct* to a one which was completely accurate. Thus, if a suspect had brown/black eyes according to the official police record, witnesses scored wrong if they described the eyes as blue, partially correct if they described them as dark, and correct if they described them as brown or black. Likewise, if the police file described the appearance of an offender as, for instance, "Northern African," a description of a witness was coded wrong if the robbers was described as Indonesian, Surinamese, or "dark type," partially wrong if the witness described the robber as White, and we coded descriptions given as Southern European and of course Northern African as correct.

For age and height the variety of description provided was far greater than for any other variable. The age could be described with a specific number (e.g., 24), a numerically bounded age range (e.g., between 19 and 22), an inferentially bounded age range (e.g., in his early 30s), an open age range (e.g., older than 20), or a nonnumerical age range (e.g., middle-aged). With such diversity, we recoded these descriptions using the rule that the more accurate the description, the smaller the number of suspects the police should have to consider in order to find the correct one. By using this rule, the accuracy of an age estimate (and *mutatis mutandis* the height estimate) was measured in terms of the size of the range of ages of suspects that the police would have to consider in order to include the age of the culprit, starting with the range initially given by the witness and expanding it

in both directions by equal amounts if necessary until this target value had been included (see Table 1 for an example).

In order to determine the size of a range for those height and age descriptions which included inferred boundaries (e.g., about 20, or late teens), a set of intuitively plausible and internally consistent numerical boundaries were determined. In addition, upper and lower limits were specified so that the scores for the 'open' descriptions could be determined. With the exception of the minimum age, these maxima and minima were those of our sample (which we assumed to be representative, in terms of age and height, of the population of robbers of commercial and residential targets in The Netherlands). The minimum age was set at 12 years, this being the minimum age of offenders recorded by the police. Our limits, then, were 12 and 55 years, and 157 cm and 200 cm.

Using these rules, the accuracy scores thus obtained were continuous and ranged from 1 to 39 for age and from 1 to 41 for height, where a lower indicated greater accuracy. These were transformed into continuous variables ranging from 1 (*least accuracy*) to 3 (*greatest accuracy*) to align them with the scale used for the other descriptors. To facilitate assessments of whole description accuracy (see below) these age and height variables were then reduced to three discrete categories by rounding the continuous values to the nearest whole number.

The accuracy measures for the individual descriptors were combined into two separate measures of accuracy. The first one, denoted accuracy of description, was a measure of the proportion of the total amount of information provided by the witness that was correct. For each descriptor either a 2 (*correct*), a 1 (*partially correct*), or a 0 (*wrong*) was given. The variable accuracy of descriptions was simply the sum of these divided by the number of descriptors given by the witness.

The measure of accuracy did not take into account the level of misinformation contained in each descriptor which could lead the police down false trails in their search for a suspect. Thus it is not possible from the accuracy of the description to distinguish between a description containing  $n$  descriptive units which are partially accurate (so they usefully limit the police's search for the suspect a little, but not as much as might be possible) and a description in which an equal number of completely accurate and completely inaccurate information units are contained. Both would have an accuracy score of 50%. Nevertheless, the latter description could be viewed as being essentially useless, since it broadens the search space as much as it limits it. Therefore, in a second measure wrong information received a negative (-2), rather than a zero weight, and partial correct information was given a weight of -1. This resulted in a variable, denoted usefulness of information, which was the sum of the separate scores of the descriptors provided by the witnesses.

## RESULTS

### Completeness of Descriptions

By far the most frequently mentioned characteristic was the sex of the suspect, with the second most prevalent characteristic being the height of the suspect (see

Table 2. Frequency of Occurrence of 24 Permanent Characteristics and Accuracy of Witnesses

Descriptor	Mentioned by witnesses (N = 2,299)		Accuracy of witnesses (%)			
	Count	Percentage	N	Correct	Partially correct	Wrong
Sex	2,175	95	1581	100	—	0
Height	1,617	70	1072	52	45	3
Appearance (including race)	1,478	64	1007	60	20	20
Skin color	1,279	56		—	—	—
Age	1,258	55	903	61	37	2
Build/posture	1,101	48	754	39	11	49
Hair color	818	36	586	73	9	18
Type of hair	793	34	553	33	0	67
Type of speech (accents and dialects)	719	31	468	32	—	68
Handedness	277	12	161	50	41	9
Face shape	271	12	78	69	31	0
Beard	241	10	177	1	—	99
Moustache	234	10	163	3	—	97
Glasses	226	10		—	—	—
Eye color	120	5	92	36	39	25
Nose	67	3	55	35	2	64
Face color (complexion)	64	3	45	38	—	62
Mouth	45	2	33	39	—	61
Eye shape	34	2	2	100	—	0
Teeth	19	1	11	27	—	73
Earrings	18	1		—	—	—
Chin	11	0	8	38	—	62
Ear size	4	0	3	33	—	67
Ears protruding	3	0	2	50	—	50

Note: Characteristics in bottom part of table (from Handedness and below) are mentioned by <30% of the witnesses.

Tables 2 and 3). Over half of the suspect descriptions also contained information about the age, the appearance (including race), the skin color, and the type of head covering or disguise worn by the offender. All of these descriptors, with the exception of the type of head covering worn, were classified as permanent characteristics and were verifiable. In contrast, the least commonly mentioned characteristic was the material of the shirt worn by the suspect. A further 15 characteristics were mentioned in less than 5% of all of the witness descriptions. The majority of characteristics appear in less than 30% of all descriptions.

Overall, the completeness of the descriptions was rather poor. The total number of characteristics mentioned in a description ranged from 1 to 23 with a median of 8 descriptors being mentioned (inter-quartile range = 6) from a possible maximum of 43. The distribution was slightly positively skewed (skewness = 0.263, S.E. skew = 0.051) with only 200 descriptions (8.7%) including between 15 and 23 descriptors.

In general, the descriptions contained more permanent than temporary characteristics: The number of permanent characteristics mentioned in a description had a median of 5 (inter-quartile range = 5), while the number of temporary char-

Table 3. Frequency of Occurrence of 19 Temporary Characteristics ( $N = 2,299$ )

	Mentioned by witnesses	
	Count	Percentage
Coat	570	25
Coat color	503	22
Coat fabric	60	3
Jacket	749	33
Jacket color	634	28
Jacket fabric	181	8
Trousers	594	26
Trousers color	410	18
Trousers fabric	15	1
Shirt	104	5
Shirt color	87	4
Shirt fabric	2	0
Shoes	195	8
Shoes color	115	5
Shoes fabric or brand	17	1
Gloves	177	8
Hat, etc.	1,175	51
Hat color	721	31
Hat fabric	99	4

acteristics ranged from 0 to 10 and had a median of 2 (inter-quartile range = 3). Both of these medians are very low in relation to the theoretical maxima for these two variables (maximum possible for temporary was 19 and for permanent 24).

#### Accuracy and Usefulness of Descriptions

Of the 2,299 descriptions, 1,650 could be matched to specific offenders. The accuracy of reporting each of the 21 verifiable descriptors varied greatly (see Table 2). For example, the percentage of descriptions in which a mentioned characteristic was completely accurate ranged from 1% (type of beard) to 100% (sex and shape of eyes). To properly evaluate these accuracy figures, the number of descriptions on which each percentage is based needs to be taken into account. Some are based on quite small numbers.

Of the 11 descriptors whose accuracy scores were obtained from samples of 100 or more, it appeared that sex was reported with perfect accuracy. Hair color descriptions were 73% correct compared to 18% incorrect, and both age and appearance were correctly reported in approximately 60% of the descriptions in which they were mentioned. In contrast, reports of facial hair were almost always incorrect (in 97% of cases for moustache and in 99% for beard), and both the type of hair and the type of speech were reported incorrectly in approximately two thirds of the descriptions in which they were reported. Finally, 51% of descriptions containing posture information were either wholly or partially correct, and 49% of descriptions containing posture information were incorrect.

**Table 4.** The Distribution of the Accuracy of the Information Provided in 1650 Verifiable Suspect Descriptions ( $N = 1,650$ )

Range of accuracy (%)	Frequency	Percentage	Cumulative percentage
0	9	0.5	0.5
0 < $x$ < 25	7	0.5	1.0
25	8	0.5	1.5
25 < $x$ < 50	188	11.3	12.8
50	56	3.4	16.2
50 < $x$ < 75	658	39.9	56.1
75	24	1.5	57.6
75 < $x$ < 100	425	25.7	83.3
100	275	16.7	100.0

Taken together, the witnesses were correct in their descriptions more often than not. In total, the witnesses gave 7,754 verifiable descriptors. Of these, 59% (4583) were correct, 17% (1,283) were partially correct, and 24% (1,888) were wrong.

Of the 1,650 descriptions, the overall accuracy of information mentioned was generally good, with the median percentage accuracy for mentioned characteristics being 72% (inter-quartile range = 34%). In only 212 (13%) of these suspect descriptions does the accuracy of the information provided fall below 50%, while in 275 (17%) cases all of the information provided was completely accurate (see Table 4).

The descriptions in which 100% accuracy was achieved were investigated further to determine whether the witnesses who provided these descriptions, or the situations from which they evolved, were any different from those in which less than perfect accuracy scores were produced. Of the witness and situational characteristics measured in the present study, only the robbery lighting conditions were found to have a significant influence upon whether the description of an offender was perfectly or less than perfectly accurate,  $\chi^2(1, N = 1650) = 7.37, p < 0.01$ . Better lighting improved the descriptions.

Usefulness scores for descriptions ranged from -17 to +4, with the median score being -2. Only 356, i.e., 22% of all descriptions, had a usefulness score above zero, indicating that they contained more useful information than misleading information. From a further analysis of the subset of descriptions attracting negative values it proved impossible to distinguish them from more useful descriptions in terms of any of the witness or situational characteristics measured. However, these descriptions were found to be significantly more complete than the remaining descriptions.

#### Prediction of Completeness and Accuracy

The effect of the independent variables on completeness and accuracy of descriptions was assessed with either two-tailed Pearson correlations or ANOVAs. An analysis of outliers showed that exclusion of outliers did not have a significant effect

**Table 5. Significant Individual Predictors of Completeness and Accuracy of Descriptions**

Predictor	Amount of information†			Accuracy of information		
	Length of description	Number of permanent characteristics	Number of temporary characteristics	Total number of characteristics	Description accuracy	Description usefulness
Age	$r = -.05^*$	$r = -.06^*$	$r = -.05^*$	$r = -.07^{**}$		
Sex		$F(1, 2280) = 12.6^{**}$	$F(1, 2280) = 6.4^*$	$F(1, 2280) = 14.4^{**}$		
Type of witness			$F^*(2, 246) = 7.58^{**}$			
Relation to object		$F^*(2, 1096) = 5.4^{**}$	$F^*(2, 1096) = 13.3^{**}$	$F^*(2, 1094) = 3.1^*$		
Delay after robbery		$F(3, 2020) = 3.2^*$	$F(3, 2020) = 3.6^*$	$F(3, 2020) = 4.7^{**}$		
Number of descriptions provided	$r = .19^{**}$	$r = -.17^{**}$	$r = -.28^{**}$	$r = -.27^{**}$	$r = .08^{**}$	$r = -.68^{**}$
Duration					$F(2, 69) = 4.0^*$	$F(2, 69) = 5.0^{**}$
Lighting conditions			$F(1, 1151) = 19.9^{**}$			
Obstructions	$F^*(2, 111) = 12.3^{**}$	$F(2, 249) = 3.8^*$				
Distance	$r = -.09^{**}$	$r = -.18^{**}$	$r = -.12^{**}$	$r = -.18^{**}$	$r = .10^*$	
Position	$F(3, 868) = 3.6^*$	$F = 3, 868) = 7.7^{**}$	$F(3, 868) = 4.3^{**}$	$F(3, 868) = 6.6^{**}$		
Disguise		$r = -.21^{**}$		$r = -.13^{**}$		$r = -.12^{**}$
Level of threat		$r = -.06^{**}$	$r = -.08^{**}$	$r = -.08^{**}$		$r = -.05^*$

\* $p < .05$ ; \*\* $p < .01$ .

† $F^*$  denotes Brown-Forsythe Statistic.

**Table 6. Direction of Influence of Independent on Dependent Variables**

Predictor	Amount of information			Accuracy of information		
	Length of description	Number of permanent characteristics	Number of temporary characteristics	Total number of characteristics	Description accuracy	Description usefulness
Age	Young → old	Young → old	Young → old	Young → old		
Sex		Female → male	Female → male	Female → male		
Type of witness			Robbery → flight			
			Robbery → preparation			
Relation to object		Employee → client	Client → Employee	Employee → passerby		
			Client → passerby			
			Employee → passerby			
Delay after robbery (days)		0 → 3+	0 → 2+	0 → 3+		
		1 → 3+	0 → 3+	1 → 3+		
Number of descriptions provided	Few → many	Few → many	Few → many	Few → many	Many → few	Many → few
Duration					Medium → long	Medium → long
Lighting conditions			Good → poor			
Obstructions	Opaque → none	None → transparent				
	Transparent → none					
Distance	Short → long	Short → long	Short → long	Short → long	Long → short	
Position	1(best) → 4	1(best) → 4	2 → 3	1(best) → 4		
	2 → 4	2 → 4		2 → 4		
	3 → 4	3 → 4		3 → 4		
Disguise		Less → more		Less → more		Less → more
Level of threat		Less → more	Less → more	Less → more		Less → more



on the results. An exception was the delay of giving a description after the robbery. There were so few descriptions given after more than 2 days after the robbery that the delay variable was recoded to four levels (0, 1, 2, and 3+ days). An analysis of the ordinal and nominal independent variables using Levene's test revealed that some the variances were significantly heterogeneous (compare Howell, 1992; Rosenthal & Rosnow, 1991). Although the ANOVA procedure seems to be remarkably robust (Winer, Brown, & Michels, 1991), the Brown-Forsythe statistic (denoted by  $F^*$ ) was used for the heterogeneous variables (Keppel, Sausley, & Tokunaga, 1992; Myers & Well, 1991; Tomarken & Serlin, 1986). The results are presented in Table 5. Where  $F$  values were observed to be significant, post hoc comparisons between all pairs of means were conducted using a test proposed by Games and Howell (1976; cited by Howell, 1992). The results of these tests are shown in Tables 5 and 6.

The effect of many of the variables on completeness was remarkably consistent across completeness measures. A shorter delay between the crime and the provision of the description, and having a shorter distance between the witness and the robber, were associated with more complete descriptions (however completeness was measured), while the duration of the incident did not have a statistically significant relationship with any of these measures. Female witnesses and those experiencing lower levels of threat were associated with descriptions including more permanent and more temporary characteristics (and therefore also more total characteristics), although these factors did not influence the length of these descriptions. Descriptions tended to be more complete where witnesses were in better viewing positions relative to the offender and when they were younger, although these factors did not affect the number of temporary characteristics included. Conversely both lighting conditions and the type of witness only influenced completeness if measured by the number of temporary characteristics mentioned. Finally, the influence of both the relationship of the witness to the target and the obstructions between the witness and the offender had no consistent relationship to the measures of description completeness.

Very few of the measured independent variables had a significant effect on the accuracy of a description. Where statistically significant results did occur, they were intuitively surprising. Given the few statistically significant relationships among the many calculated, it should be concluded that there is no association between the independent variables and both accuracy and usefulness of descriptions.

Although characteristics of the witness and the situation do not strongly predict accuracy of descriptions, the completeness of the description itself might. Correlations between the measures of completeness and accuracy (see Table 7) are all significant, but none is particularly large: at most 15% of the variance in the accuracy of descriptions was accounted for by the completeness of descriptions. The correlations between usefulness and the measures of completeness are higher, but it should be noted that these measures are not independent. More important, all correlations in Table 7 are negative, which suggests that the levels of both accuracy and usefulness decrease if witnesses provide more descriptors of the offender.

We applied various multivariate statistical techniques to assess whether a combination of the independent variables discussed above might give a better prediction of

Table 7. Pearson Correlation Coefficients for Dependent Variable Interrelationships

	Accuracy	Usefulness
Length of descriptions (characters)	-.23	-.32
Number of permanent characteristics	-.39	-.59
Number of temporary characteristics	-.21	-.25
Total number of characteristics	-.38	-.54

Note: For all correlations,  $p < .01$ .

either completeness or accuracy. Since the relationships are all quite weak, multivariate techniques did not provide any extra information above the relations presented above.

## DISCUSSION

### Major Findings

The present study of the accuracy of witness descriptions confirms the results commonly found in both laboratory and real-life studies: descriptions are usually vague and general; witnesses mostly describe general characteristics of the offenders, such as gender, race, height, and age. In describing these characteristics the witnesses are more often correct than wrong.

In evaluating these results, differences should be noted in the methods used in laboratory studies to assess description accuracy and what happens in police practice and necessarily in archival studies. In experimental research description accuracy is measured by including both inaccuracies in details reported by the witness *and* errors of omission, i.e., information available in the witnessed event which was *not* reported by the witness. In archival research such as this study errors of omission are unknown and therefore cannot be included in accuracy assessments. In police practice errors of omission of one witness can often be compensated by using descriptions of other witnesses, such that together multiple witnesses produce a more complete description of the perpetrator. The present study shows the accuracy of descriptions in such a forensic setting in terms of the information that they *do* contain. The picture that has emerged from this study, then, is that the initial descriptions elicited from witnesses to robberies are generally vague, but on the general features of the perpetrator are remarkably accurate.

### Assumptions

We argued above that conclusions on accuracy of witness descriptions from archival studies depend on two assumptions: that the individual convicted indeed committed the robbery and that the appearance of the robbers did not change significantly between the time of the robbery and the time of the arrest, when the appearance is recorded by the police.

The first assumption requires that few errors are made in criminal procedures. Someone is convicted after a three-stage (or in The Netherlands a four-stage) procedure. The police must be satisfied with the guilt of the suspect before the case is sent to the public prosecutor. The public prosecutor must be satisfied before the case goes to court. And also the court (in The Netherlands for robberies without exception a three-judge panel) must be satisfied of guilt before a conviction. And then there is still a possibility of appeal, in which in The Netherlands the case is heard *de novo*. Of course, even such a procedure can produce errors. There have been many studies of wrongful and dubious convictions (e.g., Bedau & Radelet, 1987; Borchard, 1932; Brandon & Davies, 1973; Crombag, Van Koppen, & Wagenaar, 1994; Frank & Frank, 1957; Gardner, 1952; Gross, 1987; Hale, 1961; Hannema, 1964; Hill, Young, & Sergeant, 1985; Huff, Rattner, & Sagarin, 1986; Peters, 1970, 1972; Radelet, Bedau, & Putnam, 1992; Radin, 1964; Rattner, 1988; Tullock, 1994; Wagenaar, Van Koppen, & Crombag, 1993; Waller, 1989; Woffinden, 1987; Young & Hill, 1983). Still, these known cases constitute a minute fraction of all cases handled by the courts. There is no reason to believe that wrongful convictions, known or unknown, form a notable proportion of all convictions. The estimates do not surpass a few percent (e.g., Huff et al., 1996, p. 61). Apparently, the rates of wrongful convictions do not in any way surpass the error rate common to experimental studies.

A type of error that would be particularly problematic for the present type of research is the possibility that a suspect is apprehended because he or she fits the description given by witnesses and is then convicted based on the description given. We could in part check this by noting from the case files how the suspects were found by the police. There, witnesses' descriptions play a role in only 3% of the apprehensions: 1% of the suspects were found after a witness recognized a composition drawing, another 1% pointed the suspect out in a photospread, and an additional 1% were recognized when a witness encountered the robber by accident. Most robbers are found in a manner unrelated to witness description or recognition: many, for instance, are named by other robbers (27%) or are apprehended at the scene of the crime or during their escape (22%).

The second assumption necessary to draw conclusions on the accuracy of the witness descriptions is that the offenders do not change their appearance significantly. Change of appearance is possible very quickly for some things, e.g., clothing, but more difficult for others. In the category we denoted permanent, some things are almost impossible to change, i.e., age and height, some would require plastic surgery, and others could be achieved with varying degrees of difficulty: offenders can lose or gain weight, or cut their hair. Most offenders, however, are apprehended within a reasonable period of time after their offense. In the present study, 21% were caught the same day, an additional 21% within 1 week, and another 24% within 1 month after the robbery.

If the delay between the observation by the witnesses and the recording of the real appearances of the robbers would be of importance, a negative correlation would be expected between these two variables. In the present study, total accuracy and the logarithm (because of the skewedness of the variable) of the number of days the suspect was arrested after the robbery correlated .14 ( $N = 1,223$ ,  $p <$

.01), that is, with a positive sign instead of the expected negative sign and indicating only low mutual variance.

In summary, practice and some data from the present study suggest that neither of the two assumptions necessary for drawing conclusions on the accuracy of witness descriptions holds perfectly. However, deviations from the assumptions are slight and allow for conclusions on the accuracy of witness descriptions.

### Comparison to Other Studies

Witnesses do not tell the police very much. The descriptions in our study contained a slightly lower average number of descriptors than reported by either Sporer (1992) or Tollestrup et al. (1994), and fewer permanent and clothing descriptors than Tollestrup et al.'s sample. This may be due to the inclusion in the present study of witnesses who did not see the robbery itself, but the preparations or the fleeing of the robbers. In particular, witnesses to the preparations of the robbery are comparable to both Tollestrup's fraud-witnesses and Sporer's so-called nonwitnesses. This suggests that preparation witnesses will give significantly poorer descriptions than other witnesses, and therefore supports the hypothesis that involvement in a crime significantly influences description completeness. However, the two estimates of witness involvement used in the present study—type of witness and relationship of the witness to the target of the robbery—related neither consistently nor even always significantly to description completeness. Nevertheless, such a finding does not necessarily contradict the hypothesis that different levels of involvement in a crime influence subsequent ability to provide a description, as the measures available in the present data may not accurately reflect witness involvement. For example, from the available data it was not possible to distinguish between witnesses who were victims of robbery and those who were not.

Despite the overall lower performance, the pattern of reporting clothing and other physical attributes of an observed offender was the same as that found by Tollestrup et al. The pattern and accuracy of reporting individual characteristics also largely mirrored the findings of previous research (e.g., Farrington & Lambert, 1993; Sporer, 1992). General physical attributes (e.g., sex, race, height) occurred more frequently than facial characteristics; face shape, skin color and hair details were included in descriptions more often than other facial attributes; and eye color (but not eye shape) was reported more often than lower facial features.

Overall, the picture that seems to emerge from the present study and previous research with forensic witnesses is that there is considerable concordance in forensic witnesses' ability to provide descriptions, at least in cases of robbery: The patterns of reporting particular attributes of an offender, description completeness, and the overall accuracy of information contained within descriptions are remarkably consistent, with only the accuracy of individual characteristics appearing to be somewhat more arbitrary. A detailed investigation into conditions under which certain individual attributes would be remembered accurately by a witness is needed to explain this irregularity in research findings.

### Predicting Completeness and Accuracy

To balance against this, it seems that very little information can indicate to the police when a description provided is going to be more or less accurate. The present study did find significant relationships between all measures of description completeness and description accuracy, but these relationships are very weak. As such, completeness of a description does not seem to provide a useful heuristic that could be utilized by law enforcement agencies for judging the accuracy of a description, as was also found by Wells (1985). The results of the present study suggest that the accuracy of offender descriptions cannot be predicted well from the witness and situational factors measured either.

In terms of predicting completeness, the number of descriptions provided, the distance between the offender and the witness, duration of the crime, delay, position of the witness in relation to the offender, and estimated threat experienced all had the expected effects. The use of a disguise produced the predicted detrimental effect on completeness only for permanent features and the total number of details included in descriptions, although, as all types of disguise coded in the present data set involved head covering only, it is not surprising that reporting of clothing was unaffected. Obstructions between the witness and the perpetrator failed to show a significant influence on the overall number of details included in a description, nor on the amount of clothing information reported. Unexpectedly, witnesses with no obstructions between themselves and the offender included significantly more permanent details in their descriptions than witnesses separated from the robber by transparent objects (usually glass). Two potential explanations for this can be suggested. First, it is possible that reflections on the glass (e.g., from sunlight) obscured the features of someone on the other side, and second since it is likely that most of the witnesses separated from the robber by glass were bank tellers, and also that many of these are likely to have been the direct 'victims' of the crime, it is possible that this caused the observed effect. However, with either of these explanations it is unclear why clothing and total characteristics were not similarly affected. The lighting conditions at the time of the crime surprisingly only had the expected effect on temporary characteristics included in descriptions. It is predicted that this difference will reflect a decrease in the number of clothing color details reported by witnesses to crimes at night, due to the increased difficulty in distinguishing between colors in the dark as compared to daylight. In line with the trend observed by Sporer (1992), female witnesses in the present sample included significantly more units of information of all types in their descriptions; however, interpretation of such a finding should be undertaken carefully: Sex of a witness has been found to interact with the degree of violence involved in a crime (Clifford & Scott, 1978; Macleod & Shepherd, 1986), which may underlie this main effect. Similarly, the fact that almost all of the robbers were male may have had some influence on the observed sex difference. Finally, the age of the witness was unexpectedly found to influence both the permanent and total number of characteristics reported. The only explanation that can be suggested for this effect is that it represents some confound in the data (again, for example, it is possible robbery victims tended to be young).

Perhaps the most important point to make about the present results, though, concerns the generally low correlation coefficients and  $F$  values that were observed. These suggest that little of the variance in completeness or accuracy scores can be predicted from the factors measured. Although it is possible that the measurements of predictors used may be too obtuse or inaccurate, or simply that accuracy and completeness are arbitrary and cannot be predicted, it is also possible that the present study has not measured or has been unable to measure those factors which are most influential in determining witness description accuracy. For example, although there is a standard format for interviewing witnesses which law enforcement officers are advised to follow (Projectgroep Recherche Basis Cursus, 1989), this is not necessarily always adhered to, and may vary according to the experience of the officer taking the description or the circumstances under which a statement is elicited. As questioning style affects both the amount and the accuracy of information elicited (e.g., Flin & Spencer, 1995), it is particularly unfortunate that this information was unavailable for the descriptions analyzed.

Similarly, racial differences between the witness and the offender may be an important source of variance in completeness or accuracy scores. Race of the witness was, in line with Dutch thinking on this issue, not recorded in the archive data.

#### The Police and Other Witnesses

Another potentially pertinent factor which could not be derived from the available data was the number of other witnesses that any single witness knew had seen the same crime. The potential importance of this information can be extrapolated from an effect observed in studies of altruism known as "diffusion of responsibility." This posits that a person may often fail to act when necessary if they know that others are around because they place the responsibility for this action on these others (e.g., Brehm & Kassir, 1990). It can therefore be predicted that when there are a number of witnesses to a crime they may each produce less complete and less accurate descriptions, relying on the other witnesses to report offender information. This effect, however, can also be caused by limits in time police can invest in each crime. If there are many witnesses, they share the police's time; if there are only a few, each of them becomes more important and the police may invest more time in each of them individually and thus extract longer descriptions. This post hoc hypothesis was partly supported by the data: the number of witnesses correlated  $-.20$  ( $p < .001$ ) with the number of permanent characteristics in the descriptions, but failed to correlate significantly with the other measures of completeness of description.

The data of this study show that police efforts can have another effect on the accuracy of the descriptions by witnesses. Both accuracy and usefulness of descriptions correlated negatively with all measures of completeness of descriptions, although it should be noted that since the usefulness measure was *not* divided by the number of descriptors that were provided by the witness, completeness and usefulness in part share variance because of the manner in which these measures were

calculated. The negative signs of the correlations, however, indicate that the more descriptors witnesses give of the robbers, the *less* accurate and useful their descriptions become. One explanation for this result can be drawn from previous research, which consistently shows that witnesses make few errors in free recall, but their error rate increases when pressed with questioning (Cherryman & Bull, 1996; Cutshall & Yuille, 1989; Gudjonsson, 1994; Köhnken, 1989, 1995). The present findings, then, support the conclusion that having few witnesses to a crime makes the police press for more information, and pressing for more information in turn causes descriptions to be less accurate and less useful to the police.

Further potentially important factors which could not be assessed from the available data were whether the witness was intoxicated when the crime occurred, or was wearing glasses or contact lenses if necessary. Finally it would have been useful to investigate whether there is such a thing as a 'robber stereotype' in terms of appearance information, and if so, how and to what extent it may influence information contained within witness descriptions. Holst and Pezdek (1992) demonstrated the existence of a common script for the actions that take place during a typical robbery; thus it seems plausible to expect that a stereotype also exists for the robbers themselves. Likewise, Lindsay and his colleagues drew attention to so-called default values in witness statements (Lindsay et al., 1994). For a study on robbers, such default values might mean that witnesses expect robbers to be male, and indeed almost all are male. If witnesses then report a robber to be male, we do not know whether the witness says so because he or she saw a male or just expected the robbers to be male. Both, however, produce a high degree of accuracy. Unfortunately, an investigation of the possibility of stereotyping of perpetrators influencing witnesses' accounts was beyond the scope of the present study.

The present study, together with existing research with forensic witnesses, suggests that descriptions provided by witnesses to robberies contain a consistent pattern of characteristics, although the accuracy of each attribute, with the exception of offender sex, appears to be less consistent. Furthermore, the analyses presented extended previous work on individual attribute accuracy to an assessment of whole description validity, adopting measures which seemed to be most informative in an actual forensic situation. The outcome of these analyses suggests that description accuracy in these situations is generally very good. Nevertheless, as this is only a generalization, and not all descriptions contain accurate information, the ultimate goal of research concerning the validity of offender descriptions must be to try to predict when a description will be accurate and when it will not. The present study, however, failed to find any strong predictors of description accuracy from either measures of description completeness or from various witness and situation factors. Although suggestions for improvements and additions to the predictors included in the present study were made, adoption of these suggestions will still only lead us part of the way toward a solution to this issue: Potential predictors of description accuracy do not occur in isolation, but rather in combination, and, as implied above, are likely to interact with each other. Thus, although prediction of description accuracy from witness and situational factors individually is a useful first step in understanding possible influences on the validity of these descriptions, what is ultimately needed in order to try to provide law enforcement agencies with any

useful information is a model of description accuracy in which the effects of possible predictors are considered in combination.

#### ACKNOWLEDGMENTS

We thank Ronald Roesch and two anonymous reviewers for their helpful comments on an earlier draft of this paper.

#### REFERENCES

- Adams-Price, C. (1992). Eyewitness memory and aging: Predictions of accuracy in recall and person recognition. *Psychology and Aging, 7*, 602-608.
- Alper, A., Buckhout, R., Chern, S., Harwood, R., & Slomovits, M. (1976). Eyewitness identification: Accuracy of individual vs. composite recollections of a crime. *Bulletin of the Psychonomic Society, 8*, 147-149.
- Barclay, C. D., Cutting, J. E., & Kozlowski, L. T. (1978). Temporal and spatial factors in gait perception that influence gender recognition. *Perception and Psychophysics, 23*, 145-152.
- Bedau, H. A., & Radelet, M. L. (1987). Miscarriages of justice in potential capital cases. *Stanford Law Review, 40*, 21-179.
- Borchard, E. M. (1932). *Convicting the innocent: Sixty-five actual errors of criminal justice*. Garden City, NY: Doubleday.
- Bornstein, B. H. (1995). Memory processes in elderly eyewitnesses: What we know and what we don't know. *Behavioral Sciences and the Law, 13*, 337-348.
- Brandon, R., & Davies, C. (1973). *Wrongful imprisonment: Mistaken convictions and their consequences*. London: Allen and Unwin.
- Brehm, S. S., & Kassir, S. M. (1990). *Social Psychology*. Boston: Houghton Mifflin.
- Cherryman, J., & Bull, R. (1996). Investigative interviewing. In F. Leishman, B. Loveday, & S. P. Savage (Eds.), *Core issues in policing* (pp. 147-159). London: Longman.
- Christianson, S.-Å., & Hübner, B. (1993). Hands up! A study of witnesses' emotional reactions and memories associated with bank robberies. *Applied Cognitive Psychology, 7*, 365-379.
- Clifford, B. R., & Hollin, C. R. (1981). Effects of the type of incident and the number of perpetrators on eyewitness memory. *Journal of Applied Psychology, 66*, 364-370.
- Clifford, B. R., & Scott, J. (1978). Individual and situational factors in eyewitness testimony. *Journal of Applied Psychology, 63*, 352-359.
- Crombag, H. F. M., van Koppen, P. J., & Wagenaar, W. A. (1994). *Dubieuze zaken: De psychologie van strafrechtelijk bewijs [Dubious Cases: The Psychology of Criminal Evidence]* (2nd revised ed.). Amsterdam: Contact.
- Cutler, B. L., & Penrod, S. D. (1995). *Mistaken identification: The eyewitness, psychology, and the law*. Cambridge: Cambridge University Press.
- Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987). The reliability of eyewitness identifications: The role of system and estimator variables. *Law and Human Behavior, 11*, 233-258.
- Cutshall, J., & Yuille, J. C. (1989). Field studies of eyewitness memory of actual crimes. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 97-124). New York: Springer.
- Davies, G. M. (1992). Influencing public policy on eyewitnessing: Problems and possibilities. In F. Lösel, D. Bender, & T. Bliesener (Eds.), *Psychology and law: International perspectives* (pp. 265-274). Berlin: De Gruyter.
- Davies, G. M. (1995). Evidence: Psychological perspective. In R. Bull & D. Carson (Eds.), *Handbook of psychology in legal contexts* (pp. 179-192). Chichester, England: Wiley.
- Davies, G. M., Tarrant, A., & Flin, R. H. (1989). Close encounters of the witness kind: Children's memory for a simulated health inspection. *British Journal of Psychology, 80*, 415-429.
- Deffenbacher, K. A. (1983). Eyewitness accuracy and confidence: Can we infer anything about their relationship? *Law and Human Behavior, 4*, 243-260.
- Ellis, H. D., Deregowski, J. B., & Shepherd, J. W. (1975). Descriptions of white and black faces by white and black subjects. *International Journal of Psychology, 10*, 119-123.



- Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1980). The deterioration of verbal descriptions of faces over different delay intervals. *Journal of Police Science and Administration*, 8, 101-106.
- Farrington, D., & Lambert, S. (1993, October). *Predicting violence and burglary offenders from victim, witness, and offence data*. Paper presented at the First NISCALE Workshop on Criminality and Law Enforcement, 13-16 October 1993.
- Flin, R. H., & Shepherd, J. W. (1986). Tall stories: Eyewitnesses' ability to estimate height and weight characteristics. *Human Learning*, 5, 29-38.
- Flin, R. H., & Spencer, J. R. (1995). Annotation: Children as witnesses: Legal and psychological perspectives. *Journal of Child Psychology and Psychiatry*, 36, 171-189.
- Frank, J. N., & Frank, B. (1957). *Not guilty*. New York: Doubleday.
- Games, P. A., & Howell, J. F. (1976). Pairwise multiple comparison procedures with unequal n's and/or variances: A Monte Carlo study. *Journal of Educational Statistics*, 1, 113-QS.
- Gardner, E. S. (1952). *Court of last resort*. New York: Sloane.
- Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1985). Eyewitness memory enhancement in the police interview: Cognitive retrieval mnemonics versus hypnosis. *Journal of Applied Psychology*, 70, 401-412.
- Goodman, G. S., & Schwartz-Kenney, B. M. (1992). Why knowing a child's age is not enough: Influences of cognitive, social, and emotional factors on children's testimony. In H. Dent & R. Flin (Eds.), *Children as witnesses* (pp. 15-32). Chichester, England: Wiley.
- Gross, S. R. (1987). Loss of innocence: Eyewitness identification and proof of guilt. *Journal of Legal Studies*, 16, 395-453.
- Gudjonsson, G. H. (1994). Investigative interviewing: Recent developments and some fundamental issues. *International Review of Psychiatry*, 6, 237-245.
- Hale, L. (1961). *Hanged in error*. Harmondsworth, England: Penguin.
- Hannema, U. D. (1964). *De Hogerhuis-zaak [The Hogerhuis case]*. Drachten, The Netherlands: Laverman.
- Hill, P., Young, M., & Sergeant, T. (1985). *More rough justice*. Harmondsworth, England: Penguin.
- Holst, V. F., & Pezdek, K. (1992). Scripts for typical crimes and their effects on memory for eyewitness testimony. *Applied Cognitive Psychology*, 6, 573-587.
- Howell, D. C. (1992). *Statistical methods for psychology* (3rd ed.). Belmont, CA: Duxbury.
- Huff, C. R., Rattner, A., & Sagarin, E. (1986). Guilty until proven innocent: Wrongful conviction and public policy. *Crime and Delinquency*, 32, 518-544.
- Huff, C. R., Rattner, A., & Sagarin, E. (1996). *Convicted but innocent: Wrongful conviction and public policy*. Thousand Oaks, CA: Sage.
- Keppel, G., Saufley, W. H., & Tokunaga, H. (1992). *Introduction to design and analysis: A student's handbook* (2nd ed.). New York: Freeman.
- Köhnken, G. (1989). Behavioral correlates of statement credibility: Theories, paradigms and results. In H. Wegener, F. Lösel, & J. Haisch (Eds.), *Criminal behavior and the justice system: Psychological perspectives* (pp. 271-289). New York: Springer.
- Köhnken, G. (1995). Interviewing adults. In R. Bull & D. Carson (Eds.), *Handbook of psychology in legal contexts* (pp. 235-246). Chichester, England: Wiley.
- Köhnken, G., Schimossek, E., Aschermann, E., & Hofer, E. (1995). The cognitive interview and the assessment of the credibility of adults' statements. *Journal of Applied Psychology*, 80, 671-684.
- Koneci, V. J., & Ebbesen, E. B. (1986). Courtroom testimony by psychologists on eyewitness identification issues: Critical notes and reflections. *Law and Human Behavior*, 10, 117-126.
- Kuehn, L. L. (1974). Looking down a gun barrel: Person perception and violent crime. *Perceptual and Motor Skills*, 39, 1159-1164.
- Lindsay, R. C. L., Martin, R., & Webber, L. (1994). Default values in eyewitness descriptions: A problem for the match-to-description lineup foil selection strategy. *Law and Human Behavior*, 18, 527-541.
- Loftus, E. F. (1979). *Eyewitness testimony*. Cambridge, MA: Harvard University Press.
- Loftus, E. F., Schooler, J. W., Boone, S. M., & Klein, D. (1987). Time went by so slowly: Overestimation of event duration by males and females. *Applied Cognitive Psychology*, 1, 3-13.
- Macleod, M. D., & Shepherd, J. W. (1986). Sex differences in eyewitness reports of criminal assaults. *Medicine, Science, and the Law*, 26, 311-318.
- McCloskey, M., & Egth, H. E. (1983). Eyewitness identification: What can a psychologist tell a jury? *American Psychologist*, 38, 550-563.
- McKenna, J., Treadway, M., & McCloskey, M. E. (1992). Expert psychological testimony on eyewitness reliability: Selling psychology before its time. In P. Suedfeld & P. E. Tetlock (Eds.), *Psychology and Social Policy* (pp. 283-293). New York: Hemisphere.
- Myers, J. L., & Well, A. D. (1991). *Research design and statistical analysis*. New York: HarperCollins.

