

Fire-Setting: Age Trends and Psychometrical Diagnosis of Competency Criteria for Liability¹

Wilfried Hommers

The following studies are related to the competencies of minors relevant in German civil law. However, they are specialized in several other respects. First, as a specialized field of competencies, tort law competencies of minors have been chosen, which are only part of the larger field of developmental assumptions related to the age limit of seven years in German civil law. These are regulated by § 828 BGB (German civil code of 1900) and its legal implications. Undeutsch (1967) has summarized the Supreme Court rulings on the competency criteria for liability of minors since the beginning of this century. The two principle interfaces between psychology and law in this field are *de lege lata* and *de lege ferenda* (Wegener, 1981). The former looks at cognitive development related to issues of forensic competency evaluations of minors in tort (delictual) law suits (§ 828 II BGB), whereas the latter looks at the delictual (§ 828 I BGB) and contractual (§ 104 I BGB) law assumptions of the seven-year age limit in German civil law. Hommers (1983, 1989) has reviewed the available empirical evidence, which was only indirectly concerned with this problem, and has discussed methods for further examinations. Relevant empirical aspects of both interfaces are consequently pursued in the present approach directly, but still restricted to tort law competencies.

The Supreme Court criteria for tort law (delictual) capability of minors in German civil law rest on the preassumption that the individual child has one or both of two competencies when having reached the age of seven years: knowledge of right and wrong when doing harm and some understanding of the duty to make recompense for harm. As a consequence, an individual minor may be exempted from damages only if an expertise proves that he or she did not have the necessary capabilities at the time of the tort. On the other hand, if the expertise found that the individual child had the capability to know right and wrong, the second preassumption, understanding the duty to make recompense, will not be questionable. Instead, it is assumed that the capability to know right and wrong proves the understanding of the duty to make recompense. Thus, jurisprudence implicitly assumes that the duty to make recompense is understood earlier in cognitive development than the knowledge of right and wrong in a specific case. Aside from the developmental sequence, it is also possible to test the diagnosticity assumption of whenever knowledge of right and wrong is present, understanding the duty to make recompense is also present. However, the use of the specification "some understanding" in the Supreme Court rulings

¹ Acknowledgements: The empirical studies reported were funded by a grant from the German Research Society, Bonn (DFG Ho 920/2-2) to the author. Mr. Dipl.-Psych. K. Feld assisted with the data analyses. Mrs. Pirkner typed the manuscript and the tables.

on the recompense competency shows that no high standard is held that has to be proven for attributing the necessary cognitive ability to a child.

Both competencies have intellectual and emotional-evaluative components as the Supreme Court ruled. The former may be operationalized with respect to the appreciation of the wrong against the harmed person. Thus, scores on certain IQ subtests may be used as substitutes if more direct assessment of appreciation is not feasible. For the latter, the classical notion of discernment may be the basis for its operationalization. Since no specific assessment approaches were used for the latter, the present contribution is concerned with testing the developmental assumptions for discernment in cases of purposeful and negligent torts. According to the classical discernment concept (Waibel, 1970), the right-wrong capacity may be assessed by judgment differences between negligent and inadvertent fire-setting or by judgment differences of purposeful versus inadvertent fire-setting, other variables kept constant. These differences can be measured with ratings of single stimuli presented successively or with choices between two stimuli presented simultaneously. Moreover, the low standard of the Supreme Court on the understanding of the recompense duty allowed the extension of the classical discernment notion. Therefore, the recompense capacity may also be assessed by judgment differences and in particular with judgment differences between apology conditions and no apology conditions, other conditions kept constant.

In cases of negligent torts, two capacities of the average child of the same age as the harmdoer are relevant in addition to those two discussed above. Following from § 276 BGB (culpa of the harmdoer) and § 254 BGB (contributory negligence), knowing the dangerousness of the harmful act and performing up to these cognitive capabilities is necessary for attributing liability to a minor as well as to every person. However, here an average standard is used in which an individual child's competencies are irrelevant. Thus, psychological research on average age trends of relevant cognitive and volitive competencies are of interest. However, the Supreme Court used two concepts (Waibel, 1970) for the risk-oriented cognitions involved, one being the general duty to avoid negligent acts (relevant for § 828 BGB), the other being the concrete part related to the actually existing danger (relevant for § 276 BGB). Thus, individual competency criteria, in particular of the knowledge of right and wrong, apply also to cases of minors' negligent torts.

Finally, the special tort of arson is chosen (Canter, 1980; Prins, 1986). Children's fire-setting may be of interest for practical reasons, either because it is a behavioral problem of children and juveniles (Achenbach & Edelbrock, 1981) requiring prevention and modification (Ell, 1983; Holland, 1969), or because the size of the losses involved frequently results in civil law suits against minors, in which the capacity of the minor might be tested (Dauner, 1980; Ell, 1983). Also, there are empirical results from such expertise cases (Dauner, 1980) that indicate that the relevant cognitive abilities of the minors may develop differently across different torts. Thus, arson appeared as an ecologically valid choice.

For forensic expertise judgments, procedures for the assessment of individual differences in judgments of children are apparently of interest. But, the main problem for making forensic diagnoses is methodological: to obtain estimates of reliability and to show that the assessment of delictual capacities cannot be replaced by other measures like general intelligence. Therefore, two original empirical studies employed culpa and post-act information suitable for testing the preassumption of some understanding of the recompense duty and for individual diagnoses of the discernment of negligent and purposeful fire-setting

in comparison with inadvertent fire-setting. In these studies, choice responses and rating responses were used for data collection, and direct and indirect psychometric assessment of the measurement errors (Huber, 1973) were employed for competency diagnoses. Moreover, scores on some subtests of intelligence were determined for testing their independence from the results on discernment diagnoses.

Experiment 1

The main points of the first study were to compare the mean age trends of the culpa effect and of the post-act-behavior effect, to provide results on the reliabilities of these effects in individual assessments, to test the two implicit hypotheses about the relation between knowledge of right and wrong and understanding the duty to repair, and to examine the correlations between these individual effects and intelligence measures.

Method

Twelve illustrated stories on incidents of fire-setting by a child were used as stimuli. They were presented to four groups of male and female preschool students ($N=33$, 5- to 6-year-olds), elementary school students ($N=38$, 7- to 10-year-olds), secondary school students ($N=37$, 11- to 15-year-olds) and adults ($N=40$). The stories factorially combined two damage levels (totally burnt doghouse, partly burnt barn), two culpa levels (inadvertently, when helping to search for a wallet lost by a farmer and striking a match to cast light in a dark corner; purposefully, becoming angry with the farmer who told the child to tidy up the barn as a punishment for stealing cherries), and three levels of post-act behavior (no apology by turning one's back on the farmer, apology by shaking hands with the farmer, helping to repair the burnt building). The two culpa levels were illustrated by two pictures each, the levels of the other factors by one picture each. Each picture was accompanied by a short text that was verbalized during the instruction by the examiner and later on by the subject. During the training phase of the rating task, subjects were presented all stimuli levels of each one of the stimuli factors simultaneously (damage first, followed by post-act behavior, and then culpa), and they rated them on a good-bad rating scale consisting of 13 black (right side of the scale) and 13 white bars. These bars were 0.5 cm thick and increased in size up to 6.0 cm from the middle to the ends. Afterwards, four complete story stimuli were arranged in correct order and were rated by the subjects.

The session started with the subtest MT (Block Design) of the HAWIK-R (Tewes, 1983). Afterwards, subjects entered the training phase. After two additional complete trial stimuli, which were used for retest reliability measures of the ratings, the subtest WT (Vocabulary) of the HAWIK-R followed. Then the twelve stimuli were presented. These were given in two sequences, each to half of the subjects. Finally, the subtest AV (Comprehension) of the HAWIK-R was administered.

Results

There were three main results: the age trends of the means, the assessments of individual effect differences, and correlations between individual effect differences and IQ subtest scores. Figure 1 presents the comparison of means for preschool and elementary school children as a function of three stimulus informations: Damage with levels "burnt dog-house" or "partly burnt coach-house" (horizontal axis), post-act behavior of the minor with the levels "no apology", "apology", and "helping with the repair" (curve parameters), and the culpa of the minor with the levels "inadvertend when helping with a burse search" and "purposeful as a revenge" (left and right part of each graph). Mainly, a strong effect of apology, a small effect of culpa, and a twofold disordinal age trend of apology and help with damage in the purposeful-level are visible. The means of the two older age groups resembled the 7- to 10-year-olds fairly well.

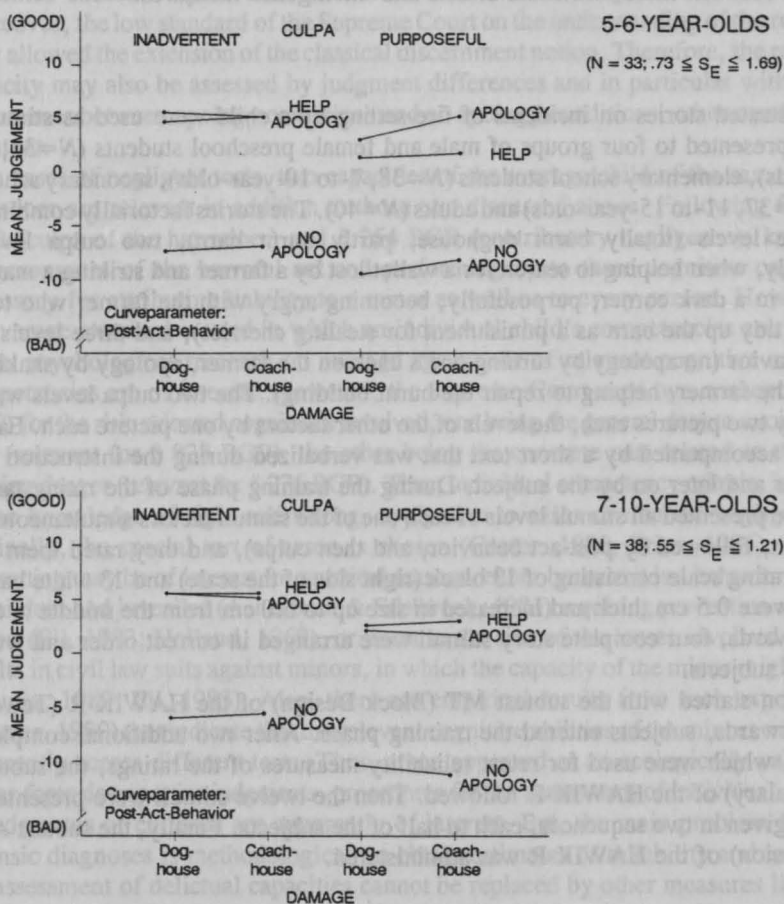


Figure 1: Means of Good-Bad-Judgements of two Age Groups on a Fire-Setting of a Minor as a Function of three Stimulus Informations.

The clearly visible main effect of culpa and of the difference between no apology and apology will be discussed with other results later (Table 1). Primarily, Figure 1 shows the results of the group means related to the 7-year age limit. As one can see, there are two "disordinal" effects in the preschoolers' means. To apologize for a purposeful fire-setting appeared better to the preschoolers than to help repair the damage. Similarly, to apologize for the larger damage appeared in their judgment better than to apologize for the smaller damage. This latter result was supported statistically by a significant triple age x damage x culpa interaction, $F(1, 144) = 4.178, p = .007$, and in part by a significant age x damage interaction, $F(3, 144) = 6.021, p < .001$. These two disordinal age trends around the age of seven years may indicate that preschoolers judge from a different recompense morality than elementary schoolers. The former may take into consideration the subjective efforts that the recompense implicates for the harmdoer, whereas the latter may center on the effect to the victim of the harm. Since similar disordinal age trends were found with various forms of another scenario (Hommers, 1986a, 1986b) and when comparing homogeneous age groups of 6- and 8-year-olds, this reliable result on the development of the understanding of the duty to make recompense seems to support the seven-year civil law age limit. However, this support exists aside from the presence of understanding the duty to make recompense up to the standard of "some understanding", as shown by results discussed below.

The mean data structure of secondary school children and adults was very similar to the elementary school children in Figure 1, except that the difference between repair help and apology was larger (see also the third column of Table 1) and their damage effects (which took the expected direction) were significant ($p < .05$ or $< .001$) in both older groups. Also, there was no main effect of age in contrast to Irving and Siegal (1983) who found a decrease of punishment as a function of age (4 groups from 7 to 17) with an arson scenario. Thus, no moral evaluation, but only punishment differences may exist in arson cases across age groups. However, age did influence the effect sizes of the stimulus components, as shown in Table 1 on the basis of individual effect sizes.

Table 1: Retest-Reliabilities (r_{11}), Cronbach's α , Group Means (M) and their Standard Deviation (SD) of the Averaged Local Effect Sizes of the Individual Subjects for Four Groups.

	N	r_{11}	Culpa (6 "Items")			Apology (4 "Items")			Repair-help (4 "Items")		
			α	M	SD	α	M	SD	α	M	SD
Preschool	33	.64	.50	-2.5	4.3	.78	12.6	7.65	-.35	-1.1	3.44
Elementary	38	.95	.78	-3.4	4.03	.94	12.4	7.67	-.01	0.7	2.39
Secondary	37	.95	.88	-6.3	4.71	.88	9.4	5.26	.27	1.9	2.19
Adults	40	.92	.81	-5.4	3.28	.91	8.2	5.81	.40	1.6	1.99
Total	148	.89	.77			.88			.11		

Because there were no further disordinal interactions, rating differences for apology and no apology or for helping with the repair and apology as well as the differences for inadvertent and purposeful fire-setting were determined for the assessment of the delictual capacities which was eventually based on the reliability of the differences estimated with Cronbach's α . As the comparison of the α -coefficients for individual effect sizes and the retest-reliabilities for judgments (first column of Table 1) show, ratings and effect sizes had reliabilities comparable to psychometric personality tests. Table 1 also presents the mean age trends in the three individual effect sizes and the relevant data for the standard error of measurement derived from the reliability as estimated from the α -coefficient and the standard deviations for the averaged individual differences of each group.

The culpa effect increased with age, $F(3, 144) = 6.452, p < .001$, although it was already significant in preschool students ($p < .01$), as seen in the lefthand side of Table 1, where standard deviations of the four age groups and α -coefficients for individual culpa effects are given. Whereas the former decreased, the latter increased with age, roughly speaking.

There was already a large effect of apology in preschool students, as expected from previous results with other scenarios of property damage (Leon, 1982), theft, personal injuries, and smearing stamps of a peer's collection (Hommers, 1988a, 1988b).

However, the stability of the effect of apology across age reported for other scenarios was not found in the arson scenario. Instead, the effect of apology on judgments decreased from preschool students to adults, $F(3, 144) = 3.965, p = .009$, as can be seen in the center of Table 1. This result was replicated in several independent studies not reported here, which varied some of the methodological features of this study by employing the arson scenario. In part, this scenario specificity in the mean age trend of the apology effect may be interpretable from the relations between victim and harmdoer or from the severity of the harm. In the arson scenario, the victim was an adult, whereas in the previously used scenarios, the victim was a peer of the harmdoer. Thus, adult subjects may identify with the victim and regard apology to be insufficient for arson.

There was also clear evidence that there is a development of the capability to understand the duty of recompense. On the one hand, there were the disordinal effects discussed above. On the other hand, there was an age-correlated increase in the distinctions made between helping to repair the damage on the one side and apology on the other, $F(3, 144) = 9.854, p < .001$, as seen in the righthand side of Table 1. According to the Supreme Court standard, this development is irrelevant for both purposes (*de lege lata* and *de lege ferenda*) of the civil law competency of understanding the duty to repair harm. But, it is in line with the low standard for understanding the duty to make recompense. Interestingly, the reliability data for these individual effects showed that there is no psychometrical basis for interpreting individual effect sizes of the difference between helping with repair and apology, although the reliability and standard deviations of individual effect sizes show the same age trend as in the two other ones.

Despite the arson-scenario-specific decreasing age trend of apology, the results clearly support the prior finding that preschool children already have some understanding of the duty of recompense, since they showed such a strong consideration of apology information (Hommers, 1988a, 1988b) or of harmdoer-compensation information (Hommers, 1986a, 1986b). As the reliability estimates and standard deviations for individual apology effects showed, however, this conclusion can be drawn individually on a psychometric basis. Thus, the results did speak directly and positively to the legal assumptions of cognitive

development. In contrast, individual culpa effects were frequently nonsignificant in preschool students, as indicated by the mean, standard deviation, and reliability in the lefthand side of Table 1. Thus, the Supreme Court's developmental hypothesis concerning the developmental relation of the two cognitions appears valid.

Individual diagnoses of the two competencies require the determination of the error level of a wrong diagnosis. The following argument leads to the choice of a high error probability: First, the burden of proof is on the harming minor. Thus, already weak signs of the presence of competency should speak against him or her. Second, one can use the probability of the adults for the randomly produced individual effect size of one, that is, literally speaking, "making an average distinction", as index for the error probability. Since the adult probabilities were $p = .24$ for culpa effects and $p = .28$ for apology effects (the error values for this cut-off score in the other groups were, in order of age, $p = .37$, $p = .30$, $p = .27$ for culpa and $p = .39$, $p = .30$, $p = .29$ for apology), it turned out that application of the adult error level of $p = .24$ would set the cut-off score for the diagnosis of knowledge of right and wrong to 2.14, 1.32, or 1.14; and for the duty to make recompense to 2.51, 1.32, or 1.28, in terms of averaged individual judgment differences of scale points for the three age groups respectively.

Another advantage of the reliability estimates for individual effect sizes is that they may be applicable to the ratings of the instruction phase. Those may present the original state of the individual better, whereas the ratings of the three-factorial stimuli may be overtrained. However, in the instruction, the method of contrasting the levels in the presentation was used, which may not be ecologically valid for the one-act-related competency diagnosis in a law suit. Applying the above cut-off score criteria showed that 11 preschool and 3 elementary school students were not showing knowledge of right and wrong in the instruction phase (3 preschool and 1 elementary school student for duty to make recompense). In the main phase of the investigation, there were 17 out of 33, 15 out of 38, 4 out of 37, and 4 out of 40 subjects (in order of age) below those age-specific cut-off scores for individual culpa effects; and only 4, 0, 1, and 3 for individual apology effects. Thus, the contrasting procedure used in the instruction phase may have made just the diagnoses of knowledge of right and wrong more likely, thereby supporting the implicit theory of the Supreme Court that at least one capacity is developed well enough at the age of seven.

The psychometrically based diagnoses of knowledge of right and wrong and of understanding the duty to make recompense were cross-classified to test the legal assumption that existing knowledge of right and wrong proves the understanding of the duty to make recompense. For comparison, this was done with two one-sided error levels ($p < .05$ and $p < .20$) for the wrong diagnoses. As the frequencies in Table 2 show, there were very few cases, five (with $p < .05$) or two (with $p < .20$) in total (all except one of them were in the adult group), in which understanding the duty to make recompense was present as a diagnosis and the knowledge of right and wrong diagnosis was not present. In contrast, 41 ($p < .05$) or 58 of the children and 30 ($p < .05$) or 33 of the adults had both diagnoses. Statistically the comparison of these frequencies (test of the hypothesis) with the two remaining cross-tabulations of the three minor groups with an absence of knowledge of right and wrong (conditional base rates) was significant ($\chi^2 = 7.96$ for the 5%-level and 5.04 for the 20%-level, $df = 1$, $p < .05$). Thus, the validity of the implicit developmental theory that understanding of the duty to make recompense is developed ahead of the

knowledge of right and wrong is supplemented by the validity of the hypothesis on the strong conditional association of the individual diagnoses.

Table 2: Cross Tabulated Psychometrical ($p < .05$ and $p < .20$ for the Numbers in Parentheses) Diagnoses of Right-Wrong-Knowledge and of Understanding the Repair-Duty on the Basis of Quantitative Judgments of Three-Factorial Descriptions of a Minor's Fire-Setting for Children and Adults.

		Children		Adults	
		Right-wrong-knowledge present			
		Yes	No	Yes	No
Understanding the recompense-duty present	Yes	41 (58)	53 (46)	30 (33)	5 (4)
	No	1 (0)	13 (4)	4 (0)	1 (1)

Finally, there were only two (out of 27) significant ($p < .05$) correlations between the three HAWIK-R scores and the three individual effect measures in the three groups. Only one of them was in the preschool student group and might have indicated that, at this age, some association between vocabulary and the knowledge of right and wrong as measured by the individual culpa effect may exist ($r = .45$). The other significant correlation was in the secondary school student group between block design and the culpa effects, and had no relevance for the legal questions since it was negative ($r = -.37$). The results on the multiple regression analyses involving sex and age additionally showed that the individual effect assessments could not be replaced by the intelligence measures, although there were statistically significant overall associations as indicated by the multiple regression coefficients for the three criteria variables ($p < .001$, $p < .05$, and $p < .001$). Thus, applying confirmative statistical standards, the correlations of judgments and intelligence showed that intelligence does not allow a prediction of reliable and valid individual differences in the moral rating task. Thus, the evaluative capacities of liability competency may not be replaced with intelligence testing.

Experiment 2

The main point of the second study was to involve the diagnosis of understanding the duty to avoid negligent acts as a topic of the procedure. This means that an interpretation of this moral knowledge is made that generalizes the discernment approach of Experiment 1 to the assessment of the evaluative component of knowledge of right and wrong of negligence. Thus, ratings on a negligent fire-setting were added.

Method

The picture for the negligent fire-setting showed a child making a fire near the barn, when the accompanying text added the information that the child had recently been told not to do this. This stimulus was presented as a third level of the culpa stimuli factor in addition to the levels of inadvertent and purposeful fire-setting. However, instead of

the apology effect, two new aspects were included: Another kind of stimulus information was combined with the culpa levels, and the influence of methodological variations of the individual assessment per contrasting procedure was investigated. First, the stimuli were factorial combinations of the three culpa levels mentioned before with the two levels of the post-act event factor, which had a third-party compensation level (the picture showed a farmer in an insurance office receiving money) and the apology level of Experiment 1. Second, two judgment procedures were used instead of the single rating procedure in Experiment 1.

On the one side, subjects had to give two ratings during the assessment task. Thus, a direct assessment of the standard error of measurement could be used by a repeated measurement approach aside from the indirect consistency approach to reliability estimates of Cronbach's α (indirect assessment of the standard error of measurement) used in Experiment 1. On the other side, subjects gave their ratings after six paired comparisons of culpa levels with constant post-act event conditions. All three paired comparisons with the three apology stimulus combinations and all three paired comparisons with the three third-party-compensation stimulus combinations were given. Thus, two ratings were available for each of the six stimulus combinations and each pair of culpa levels was used twice. During the training phase, subjects were presented all of the stimuli levels of each one of the stimuli factors simultaneously (the two damage levels of Experiment 1 first, followed by the two post-act events, and then the three culpa levels) and rated them on the good-bad rating scale as in Experiment 1. Additionally, there were three pretrials of paired comparison, in which subjects first arranged the two correct story sequences and then indicated which of the two story events was better. Half of the subjects proceeded with the high damage level, the other half with the low damage level. This produced no effects on the results. After the main phase, a simplified procedure was run in a third phase to check its effect on the results. A damage level was shown and afterwards the three culpa levels were added verbally and subjects had to rate the combination of visible and auditory stimuli each. Subjects were twenty-five 5- to 6-year-olds, twenty-three 7-year-olds, and thirty-nine 8- to 9-year-olds, as well as 49 parents of the children. For adults, the procedure was shortened to a questionnaire format in which only ratings had to be made.

Three IQ subtests of the AID (Kubinger & Wurst, 1985) were given to each child similar to Experiment 1. Two of them were analog to the AV- and MT-subtests of the HAWIK-R used in Experiment 1, the third was the BO equivalent (Picture Arrangement). Since correlations and multiple regression analyses yielded no contradictory results to those of Experiment 1 (instead even nonsignificant multiple regression coefficients), these results are not discussed further in the result section of Experiment 2. However, the replicated absence of the correlations supported the conclusion that intellectual and emotional-evaluative components of the delictual capacity criteria have to be assessed with different methods.

Results

The mean age trends showed no main effect of age for the doghouse condition, but adults rated more harshly in the coachhouse condition, $F(1,70)=7.90$, $p=.006$. Both results once more contradicted Irving and Siegal (1983), who found a decrease in punishment with age. Also, the culpa effect was stronger in adults with the damage condition coachhouse, $F(2,140)=3.37$, $p=.037$. Within the groups of children, age had no effect

on mean judgments of culpa levels. Furthermore, there was an age-correlated increase of the difference between ratings on the apology combinations and the third-party-compensation combinations. The conditions with third-party compensation by an insurance company were rated significantly worse than the conditions with apology only by the adults, and independently of the damage condition, $F(1, 102) = 21.86, p < .001$. This effect paralleled the increase in the distinction between apology and helping with repair found in the prior study. The means (standard deviations) of the children were $-8.8 (2.9)$, $-5.8 (4.4)$, and $1.2 (5.2)$ for purposeful, negligent, and inadvertent conditions (negative signs for the bad side of the scale), data for adults were $-9.5 (2.4)$, $-7.8 (2.9)$, and $1.4 (4.7)$. Thus, the negligent stimulus generally seemed to be definitely bad to all groups. There were clear, reliable individual differences in the evaluations when both types of measurement error were employed (see below). Also, clear age trends in the measurement error indicated that preschoolers' data were not reliable enough to make individual assessments of capacities. Thus, after taking into consideration the psychometric information, an age trend associated with the ratings of the culpa levels turned out to exist.

The analyses on the methodological aspects showed that the examined methodological aspects were irrelevant for the results. Group reliability estimates of the ratings and of the three rating differences among the culpa levels showed that children's ratings (0.76, 0.79, and 0.80 for decreasing culpa) and their rating differences (0.42, 0.76, and 0.72 for purpose-negligence, purpose-inadvertence, and negligence-inadvertence) were somewhat less reliable than adults' (0.86, 0.88, and 0.89 as well as 0.81, 0.85, and 0.88 for ratings and rating differences respectively), but high enough for the purpose-inadvertent and negligent-inadvertent differences (0.76 and 0.72) to have diagnostic impact. The standard deviations of the differences were 2.7, 5.8, and 5.5 for children and 2.2, 4.5, and 4.9 for adults. Among children, α estimates of reliability increased with age, but considerably more in the purpose-negligent differences (about 0.4) than in the two other differences (about 0.1). Similarly, across all age groups, direct assessments of the standard error of measurement decreased with age (6.9, 4.6, 2.9, and 2.3 in average, and 13.4, 10.6, 9.3, and 4.6 in the maximum individual standard error in the sequence of age). Finally, paired comparisons of children became more consistent in the choice tripels and in retests within their age range. Thus, the applicability of the procedure for the assessment of delictual competencies of minors of at least 7 years of age appeared to be psychometrically sound. In particular, the psychometrically based diagnoses of the direct and indirect assessment methods for the standard error of measurement produced less than 18% of discrepancies in the 85 children (versus 6% in adults). Also, diagnoses on the basis of ratings were the more secure diagnostic decisions, because the paired comparison procedure less frequently yielded (10%) contradictory results in cases of rating-based diagnoses (versus 36% the other way round). Comparison of the instruction phase, the main phase, and the final phase showed that in the final phase the same frequency of diagnoses for the distinction between purposeful and inadvertent (58 of 84 children) was obtained, whereas the two other diagnoses occurred slightly less often (5 to 6 depending on the standard error used). But, still, 42 children distinguished with $p < .05$ between the negligent and the inadvertent fire-setting in their judgments of the final phase. Thus, it appears that learning would not account for much of the competency diagnoses. Table 3 presents the main advantage gained from the reliability assessment in terms of revealing an age trend in the differences among the culpa levels that was not visible in the mean data and also

not visible in the arbitrarily chosen numeric difference of one scale unit (lefthand side of Table 3) used as a measure of the knowledge of right and wrong as assessed by the rating response.

Table 3: Frequencies of Distinction Capacity Diagnoses of 4 Age Groups for 3 Assessment Procedures: Numeric Comparison of Individual Judgments, Psychometric Decision on the Basis of the Individual Measurement Error (Direct), Psychometric Decision on the Basis of Group Reliability (Indirect).

	Numeric				Psychometric										
	Children		Adults		Direct ($p < .10$)				Indirect ($p < .05$)						
	5-6	7	8-9	Σ	Children		Adults		Children		Adults				
5-6	7	8-9	Σ	5-6	7	8-9	Σ	5-6	7	8-9	Σ	5-6	7	8-9	Σ
Purpose worse negligence	7	11	25	(43)	24	3	3	10	(16)	11	3	3	9	(15)	14
Elsewise	18	12	15	(45)	25	22	20	30	(72)	38	22	20	31	(72)	35
Purpose worse inadvertance	20	17	37	(74)	48	9	15	34	(58)	44	12	15	31	(58)	45
Elsewise	5	6	3	(14)	1	16	8	6	(30)	5	13	8	9	(30)	4
Negligence worse inadvertance	20	18	37	(75)	47	6	12	32	(50)	38	8	14	25	(47)	40
Elsewise	5	5	3	(13)	2	19	11	8	(38)	11	17	9	15	(41)	9

Table 3 shows that the number of individuals who can clearly ($p < .10$ as well as $p < .05$) be diagnosed to distinguish between purpose and inadvertent and between negligent and inadvertent fire-setting on psychometric grounds increased with age independently of the method of estimating the standard error of measurement. In particular, most preschoolers cannot be diagnosed as having the necessary capacities of judgment only if one of the two psychometric procedures is employed. Moreover Table 3 shows that, in contrast to the preschool students, the majority of 7-year-olds can still be diagnosed to have distinguished between inadvertent fire-setting and purposeful or negligent fire-setting, when psychometric procedures are applied to the judgments. This age trend drastically underlines another empirical part of the validity of the 7-year-age limit in the tort law of Germany, that is, the increase in the feasibility of the assessments on the individual level.

These psychometrically based results add to those of Schleifer, Shultz, and Lefebvre-Pinar (1983) who found, only reporting means of 5- and 7-year-olds, an onset of the distinction between inadvertent and voluntary as well as foreseeable damages but not between voluntary and foreseeable. Since the distinction between purposeful and negligent fire-setting was rare in children, in the reported psychometrically based diagnoses the conclusion was also substantiated that there may be no clear concept of negligence up to 10 years of age. Hook (1989), using a refined broken-cup task from Piaget (1932), claimed so merely on the basis of group means. However, employing the present methods allowed a more precise statement on an individual level. Finally, these results are in sharp contrast to

Tisak and Turiel (1984), who found no developmental trends (6 to 10 years) in several interview variables about a prudence incidence describing a child cutting the knee due to running too fast and falling down. The difference may be a consequence of several methodological dissimilarities in the approaches. But, it may indicate also that moral knowledge about negligence with respect to one's own damage develops earlier than moral knowledge about negligence with third-party damages. It is interesting that this was assumed by an Upper German Court (OLG Celle 1968, see Hommers, 1983, p. 26) with respect to a contributory negligence case in which the victim was a child. However, this cross-study conclusion would need direct research.

Discussion

To summarize the results on the two aspects for research of legal psychology in the tort law competencies:

De lege ferenda in support of the 7-year-age limit was found in the results: (a) two "disordinal" age trends in the means of preschool and elementary school students; (b) the presence of the apology effect in preschool students; and (c) in accordance with the low standard qualification "some understanding", signs of an ongoing development of that duty which is not taken to be relevant by the Supreme Court, i.e., the decrease of the rating differences between apology and no apology with age as well as the increasing distinction between the levels of apology and of helping with repair and between the levels of third-party compensation and of apology; and (d) the increase in the reliability of individual effect sizes from preschool to elementary school age.

De lege lata was relevant in the results (a) that the standard error estimates were sufficient for individual diagnoses from the age of seven years on; (b) that the psychometrically based distinction between the two or three culpa levels increased within elementary school age but was only rarely present in preschool students; (c) that the implicit Supreme Court theory on the indicativeness of knowledge of right and wrong for understanding the duty to make recompense was valid; and (d) that the diagnosis of competencies on the basis of indirect or direct reliability estimates was independent of intelligence measures and of learning within the task.

As a general consequence for forensic psychology, criticisms like those of Bresser (1988) against a trend in forensic or legal psychology that uses psychometrical, statistical, and mathematical-experimental methods seems unqualified. Apparently, such legal psychology can provide tools and results that are clearly useful for any scientifically defined de lege ferenda or de lege lata interface of psychology and law (see Steller, 1989, for another reply to Bresser's biographical position). Thus, even today, one should follow Marbe's (1913) example of introducing experimental and statistical procedures to the civil law expertise work of psychologists.

In contrast, the previous simplification strategy of basing expertise judgments on IQ measures alone (Bresser, 1972; Ell, 1983) is highly questionable. Conceptually, it misses the standards set by the Supreme Court. These are twofold: intellectual (use of adequate IQ subtests) and emotional-evaluative (use of the moral judgment task employed here). Empirically, it is in contrast to the zero correlations between the measures of the two standards. However, it is true that forensic expertise needs more complex (multi-method)

integrational diagnostic reasoning than that involved in any one single diagnostic tool, as has been outlined by Wegener and Steller (1986) and Steller (1988). In support of their well-taken broad approach to forensic expertise judgments, it appears recommendable to extend the psychometrical research approach as employed here for a first time in a meaningful manner to other forensic fields in the interface between law and psychology.

References

- Achenbach, T.M., & Edelbrock, C.S. (1981). Behavioral problems and competencies reported by parents of normal and disturbed children aged four through sixteen. *Monographs of the Society for Research in Child Development*, 46, (1, Serial No. 188).
- Bresser, P.H. (1972). Die Beurteilung der Jugendlichen und Heranwachsenden im Straf- und Zivilrecht. In H. Göppinger, & H. Witter (Eds.), *Handbuch der Forensischen Psychiatrie*, Vol. 2. (pp. 1284-1313). Berlin: Springer.
- Bresser, P.H. (1988). Über die Grenzen psychiatrischer Dokumentation: Was wird nicht abgebildet? *Forensia*, 9, 163-173.
- Canter, D. (Ed.) (1980). *Fires and human behavior*. Chichester: Wiley.
- Dauner, I. (1980). *Brandstiftung durch Kinder. Kriminologische, kinderpsychiatrische und rechtliche Aspekte*. Bern: Huber.
- Ell, E. (1983). *Wenn Kinder zündeln. Vorschläge zur Feuererziehung*. Tübingen: Katzmann.
- Holland, C.J. (1969). Elimination of fire-setting in a seven-year-old boy. *Behaviour Research and Therapy*, 7, 135-137.
- Hommers, W. (1983). *Die Entwicklungspsychologie der Delikts- und Geschäftsfähigkeit*. Göttingen: Hogrefe.
- Hommers, W. (1986a). Non-Additivität als Beleg für die moralische Natur der Integration von Schaden und Ersatzleistungen. *Archiv für Psychologie*, 138, 71-90.
- Hommers, W. (1986b). Zusammenwirken von Schaden und Ersatzleistung im moralischen Urteil. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 18, 12-21.
- Hommers, W. (1988a). Entschuldigung und Entschädigung für einen Diebstahl. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 20, 121-133.
- Hommers, W. (1988b). Die Wirkungen von Entschuldigung und Entschädigung auf Strafurteile über zwei Schadensarten. *Zeitschrift für Sozialpsychologie*, 19, 139-151.
- Hommers, W. (1989). Die Entwicklung der Einsicht in das Delikt. In S. Bäuerle (Ed.), *Kriminalität bei Schülern: Ursachen und Umfeld von Schülerkriminalität*, Vol. 1 (pp. 97-116). Stuttgart: Angewandte Psychologie.
- Hook, J.G. (1989). Heider's foreseeability level of attribution: Does it come after intentionality? *Child Development*, 60, 1212-1217.
- Huber, H.P. (1973). *Psychometrische Einzelfalldiagnostik*. Weinheim: Beltz.
- Irving, K., & Siegal, M. (1983). Mitigating circumstances in children's perceptions of criminal justice: The case of an inability to control events. *British Journal of Developmental Psychology*, 1, 179-188.
- Kubinger, K.D., & Wurst, E. (1985). *AID. Adaptives Intelligenzdiagnostikum*. Weinheim: Beltz.
- Leon, M. (1982). Rules in children's moral judgments: Integration intent, damage, and rational information. *Developmental Psychology*, 18, 835-842.
- Marbe, K. (1913). Psychologische Gutachten zum Prozeß wegen des Mülheimer Eisenbahnunglücks. *Fortschritte der Psychologie und ihrer Anwendungen*, 1, 339-374.
- Piaget, J. (1932). *Le jugement moral chez l'enfant*. Paris: Alcan.
- Prins, H. (1986). *Dangerous behavior, the law, and mental disorder*. London: Tavistock.
- Schleifer, M., Shultz, T.R., & Lefebvre-Pinard, M. (1983). Children's judgments of causality, responsibility, and punishment in cases of harm due to omission. *British Journal of Developmental Psychology*, 1, 87-97.
- Steller, M. (1988). Standards der forensisch-psychologischen Begutachtung. *Monatsschrift für Kriminologie und Strafrechtsreform*, 71, 16-27.

Steller, M. (1989). Gemälde oder Bauwerke? Anmerkungen zur Objektivierung der forensischen Begutachtung. *Monatsschrift für Kriminologie und Strafrechtsreform*, 72, 155-159.

Tewes, U. (1983). *Hamburg-Wechsler-Intelligenztest für Kinder, Revision 1983*. Bern: Huber.

Tisak, M.S., & Turiel, E. (1984). Children's conceptions of moral and prudential rules. *Child Development*, 55, 1030-1039.

Undeutsch, U. (1967). Delikthaftung junger Menschen. In U. Undeutsch (Ed.), *Forensische Psychologie* (pp. 567-597). Göttingen: Hogrefe.

Waibel, E. (1970). *Verschuldensfähigkeit*. Berlin: Duncker & Humblot.

Wegener, H. (1981). *Einführung in die Forensische Psychologie*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Wegener, H., & Steller, M. (1986). Psychologische Diagnostik vor Gericht. Methodische und ethische Probleme forensisch-psychologischer Diagnostik. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 7, 103-126.

Blatt, R. (1987). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 8, 1-10.

Blatt, R. (1988). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 9, 1-10.

Blatt, R. (1989). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 10, 1-10.

Blatt, R. (1990). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 11, 1-10.

Blatt, R. (1991). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 12, 1-10.

Blatt, R. (1992). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 13, 1-10.

Blatt, R. (1993). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 14, 1-10.

Blatt, R. (1994). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 15, 1-10.

Blatt, R. (1995). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 16, 1-10.

Blatt, R. (1996). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 17, 1-10.

Blatt, R. (1997). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 18, 1-10.

Blatt, R. (1998). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 19, 1-10.

Blatt, R. (1999). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 20, 1-10.

Blatt, R. (2000). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 21, 1-10.

Blatt, R. (2001). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 22, 1-10.

Blatt, R. (2002). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 23, 1-10.

Blatt, R. (2003). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 24, 1-10.

Blatt, R. (2004). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 25, 1-10.

Blatt, R. (2005). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 26, 1-10.

Blatt, R. (2006). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 27, 1-10.

Blatt, R. (2007). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 28, 1-10.

Blatt, R. (2008). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 29, 1-10.

Blatt, R. (2009). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 30, 1-10.

Blatt, R. (2010). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 31, 1-10.

Blatt, R. (2011). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 32, 1-10.

Blatt, R. (2012). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 33, 1-10.

Blatt, R. (2013). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 34, 1-10.

Blatt, R. (2014). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 35, 1-10.

Blatt, R. (2015). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 36, 1-10.

Blatt, R. (2016). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 37, 1-10.

Blatt, R. (2017). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 38, 1-10.

Blatt, R. (2018). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 39, 1-10.

Blatt, R. (2019). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 40, 1-10.

Blatt, R. (2020). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 41, 1-10.

Blatt, R. (2021). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 42, 1-10.

Blatt, R. (2022). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 43, 1-10.

Blatt, R. (2023). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 44, 1-10.

Blatt, R. (2024). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 45, 1-10.

Blatt, R. (2025). Die Beurteilung der Intelligenz bei Heranwachsenden im Straf- und Jugendrecht. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 46, 1-10.