

Finnish District Judges' Assessments of Live Versus Video-Mediated Party Statements in
Court

Jonas Wilkman, 1802234
Master's Thesis in Psychology
Supervisors: Julia Korkman and Jan
Antfolk
Faculty of Arts, Psychology and
Theology
Åbo Akademi, 2022

FINNISH DISTRICT JUDGES' ASSESSMENTS IN COURT

Subject: Psychology	
Author: Jonas Wilkman	
Titel: Finnish District Judges' Assessments of Live Versus Video-Mediated Party Statements in Court	
Supervisor: Julia Korkman	Supervisor: Jan Antfolk
Abstract: <p>Professionals within the judicial system sometimes believe they can assess whether someone is lying or not based on cues such as body language and emotional expression. Research has, however, shown that this is impossible. The Finnish Supreme Court has also given rulings in accordance with this demonstrated fact. There has also been previous research on whether party or witness statements are assessed differently in court depending on whether they are given live, via videoconference, or via prerecorded video. In the present study, we investigated how a Finnish sample of district judges ($N=47$) assigned probative value to different variables concerning the statement or the statement giver, such as body language and emotional expression. We also investigated the connection between the judges' beliefs about the relevance of body language and emotional expression and their preference for live statements or statements via videoconference. The judges reported assigning equal amounts of probative value to statements given live and statements given via videoconference. However, judges found it easier to detect deception live, and this preference correlated with how relevant they thought body language is when assessing the probative value of the statement. In other words, a slight bias to assess live statements more favorably than statements given via videoconference might still exist. More effort needs to be put into making judges and Supreme Courts aware of robust scientific results that have been the subject of decades of research, such as the fact that one cannot assess whether someone is lying or not based on cues such as body language.</p>	
Key words: reliability, credibility, veracity, truthfulness, assessment, deception detection, probative value, scientist-practitioner gap	
Date: 4 Jan 2022	Page count: 48
Level: Master's thesis	

Swedish abstract

Ämne: Psykologi	
Författare: Jonas Wilkman	
Titel: Finska tingsdomares bedömningar av partsutlåtanden givna på plats i rätten eller via videokonferens	
Handledare: Julia Korkman	Handledare: Jan Antfolk
<p>Abstrakt: Rättssystemets aktörer kan ha en uppfattning att de kan bedöma huruvida någon talar sanning eller inte utgående från personens kroppsspråk eller känslouttryck. Rättspsykologisk forskning har dock visat att detta är omöjligt. Högsta domstolen i Finland har även gett prejudikat i linje med detta. Tidigare forskning har vidare undersökt huruvida det finns en skillnad i hur parts- och vittnesutlåtanden bedöms beroende på om de ges fysiskt på plats i rättssalen, via videokonferens eller som videoinspelade utsagor. I vår studie undersökte vi hur ett sampel av finska tingsdomare ($N=47$) tillmäter bevisvärde åt olika variabler gällande en parts utlåtande eller partens beteende, såsom partens kroppsspråk eller känslouttryck. Vi undersökte även huruvida det finns ett samband mellan domarnas uppfattningar om hur viktiga kroppsspråk och känslouttryck är i samband med bevisvärderingen och huruvida de föredrog utlåtanden givna på plats i rättssalen eller via videokonferens. Domarna angav att de tilldelar lika mycket bevisvärde åt utlåtanden givna live och utlåtanden givna via videokonferens. De ansåg dock det vara lättare att bedöma om någon talar sanning eller inte när en part hörs på plats i rätten, och denna preferens korrelerade med hur viktigt de ansåg kroppsspråk vara i bevisvärderingen över lag. Med andra ord kan det ännu finnas en preferens för utlåtanden givna live. Mera insatser behövs för att öka medvetenheten bland domare och Högsta domstolar om sådana forskningsresultat som varit föremål för decennier av forskning, såsom det faktum att man inte kan bedöma huruvida någon talar sanning eller inte utgående från personens kroppsspråk.</p>	
Nyckelord: tillförlitlighet, trovärdighet, sanningsenlighet, bedömning, lögn detektion, bevisvärde	
Datum: 4.1.2022	Sidtal: 48
Nivå: Pro gradu -avhandling	

TABLE OF CONTENTS

Introduction.....	1
Deception Detection Amongst Legal Professionals.....	3
Case Law from the Finnish Supreme Court.....	4
Live Versus Video-Mediated Statements in Court.....	6
The Current Study.....	8
Terminological Challenges.....	8
Aims of the Study.....	9
Hypotheses.....	10
Method.....	11
Ethical Permission.....	11
Participants.....	11
Materials.....	11
Procedure.....	14
Results.....	14
Statistics and Data Analysis.....	14
Discussion.....	21
The Scientist-Case Law Gap.....	25
Future Research and Practical Implications.....	26
Limitations.....	28
Conclusions.....	29
Swedish Summary – Svensk sammanfattning.....	30
References:.....	39

Finnish District Judges' Assessments of Live Versus Video-Mediated Party Statements in Court

Research on deception and deception detection has increased in the last decades, and new methods for differentiating between truthful and deceptive accounts are constantly being devised and envisioned (Curtis, 2021; Gamer, 2014; Levine, 2014; McCornack et al., 2014). Bond and DePaulo's (2006) meta-analysis comprising 206 studies with 24,483 deception assessments showed that assessments of whether someone is telling the truth or not are correct about 54% of the time (Bond & DePaulo, 2006). In other words, this performance hardly surpasses that of a coin flip. It has moreover been argued that the around 4% that is above chance is due to identifying a small group of liars that are very easy to spot (Levine, 2010, 2014; for a summary on deception detection research, see, e.g., Landström 2008, 2012; Granhag et al., 2015; Schelin, 2006; Strömwall, 2010; Vrij 2014; Väisänen & Korkman, 2014). The legal community has also become increasingly aware that assessing deceit based on nonverbal cues is not reliable (KKO 2021:5; KKO 2019:84; KKO 2013:97; KKO 2013:96). When judges engage in deception detection this ultimately affects the outcome of the case through the legal concept of *probative value* (fi. *näyttöarvo*; sv. *bevisvärde*). The court determines what has been proven and what has not been proven in the case and shall consider the probative value of the evidence on the basis of free consideration of the evidence (Finnish Code of Judicial Procedure, Chapter 17, Section 1, in force as of 1 January 2016). In doing so, the court is required to first assess the probative value of every individual piece of evidence (e.g., a witness statement or a party statement given for probative purposes) and subsequently assess the probative value of all pieces of evidence as a whole (HE 46/2014 vp; KKO 2019:2; Rautio & Frände, 2020; Virolainen & Martikainen, 2010). In other words, when district judges engage in deception detection, it should always be done within the framework of assessing a statement's probative value.

Aside from the research focusing on nonverbal indicators of deceptions, other studies have focused on the statement itself to establish if there are any criteria-based differences between truthful and fabricated statements. Several models have already been suggested, such as the Swedish formal structure analysis by Trankell and its German counterpart statement validity assessment (SVA), the reality monitoring (RM) approach, and Sapir's Scientific Content Analysis technique. SVA has been the focus of much research, and its most salient part is the criteria-based content analysis, which is a set of 19 criteria that should be prevalent in deceitful statements. RM, with its roots in cognitive psychology, also focuses

on the differences between truthful and deceptive statements and contains a list similar to the one in criteria-based content analysis. The lists in the SVA and RM approach contain criteria such as “logical consistency”, “superfluous details”, “spontaneous corrections”, “clarity”, “reconstructability” and “realism” (see, e.g., Hirvelä, 2006; Masip et al., 2005; Schelin, 2007; Trankell, 1982; Willén & Strömwall, 2012). Many of the studies assessing the discriminatory power of these models, especially older studies, have been criticized for having methodological problems (Masip et al., 2005; Vrij, 2005). There has also been a call for more integration between the approaches, especially between the CBCA and the RM approach (Masip et al., 2005). In other words, the models have shown a discriminatory power above chance, but more research and refinement of the criteria are needed before they can be used in criminal cases. For such use, a model would need to reach a scientifically established discriminatory power that exceeds the evidential threshold of “beyond reasonable doubt” (fi: “ei jää varteenotettava epäilyä”, Rautio & Frände, 2020).

When these models have been tested, some criteria have received more empirical support than others. In the CBCA approach, the criterion “quantity of detail” has received substantial support and may confidently be used as a discriminator (Vrij, 2005). Some studies have also shown support for the criterion “clarity”, sometimes operationalized as “clarity/vividness”, while other studies do not support it as a reliable discriminator. (Masip et al., 2005). Furthermore, there is convincing research evidence of the fact that false narratives may be internally coherent and logical (Vrij, 2005). This means that the coherence of a statement is not a valid criterion that can be used to discriminate between truthful and deceptive accounts. There is also research suggesting that the consistency between statements that have been given by the same statement giver but at different times, for example, during the pre-trial investigation contra the main hearing in court, may not be a reliable discriminator between truthful and deceptive accounts (Granhag, Strömwall & Jonsson, 2003). This has been thought to be an effect of the repeat versus reconstruct hypothesis, according to which those giving deceitful accounts will try to repeat the same things they have stated previously, whereas those who tell the truth merely try to remember what happened and do not concern themselves that much with what they said previously (Granhag, Strömwall & Jonsson, 2003).

So-called “paraverbal” indicators of deception, such as length of the statement, have also been studied. In a meta-analysis by Sporer and Schwandt (2006), it was shown that message duration was slightly, but significantly, lower for deceitful statements than for

truthful ones (unweighted effect size: Cohen's $d = -.113$, weighted effect size: Cohen's $d = -.080$). This is a good example of how statistical significance alone cannot be used as a basis for practical recommendations. The effect sizes are so small as to be practically irrelevant in the court context, something which was highlighted by the researchers themselves (Sporer & Schwandt, 2006). According to Cohen (1988), an effect size like this is similar to statistically comparing the mean height of 15- and 16-year-old girls. This height difference, although observable in large groups, is insufficient and not reliable enough when trying to assess a teenage girl's age based on her height. In other words, when analyzing whether statement length differs between truthful and deceptive accounts, a slight difference can be found when analyzing large amounts of data, but in the individual case, statement length is not something the judge can use to discriminate between truthful and deceptive statements.

Deception Detection Amongst Legal Professionals

Although research strongly shows that there are no reliable cues to deception, people continue to erroneously believe that such cues exist (DePaulo et al., 2003). International studies including people with highly varied backgrounds have shown that the most common cue people look for is gaze aversion (Global Deception Research Team, 2006; Granhag & Strömwall, 2004). Research has shown that this is also thought to be the most reliable cue for detecting deception amongst practitioners within the legal field (Granhag & Strömwall, 2004). One of the reasons for why this erroneous belief persists is the lack of feedback. Practitioners seldom learn whether their assessment of deception were right or not. Judges, police officers, and other professionals tend to be more confident in their deception detection skills than lay persons, but research has shown that their deception detection is at the same level as that of laypersons (Kassin, Meissner & Norwick, 2005; Landström, 2008; Porter et al., 2000; Rosenfeld & Penrod, 2011; Schelin, 2006; Strömwall & Granhag, 2003; Vrij, 2010, 2014).

One further source of error that judges and other persons who try to discern truth from lies tend to be affected by is the Othello-Error. The Othello-Error is the tendency to see signs of nervousness as indicative of deception. Especially relevant for the courtroom scenario is the finding that a person who has no intention to lie may be just as nervous as another person who is lying (Vrij, 2014).

There is also research regarding emotional reactions to such events as sexual abuse, trauma, and stress, as well as how these reactions are displayed in the court room and subsequently affect the probative value assessment. However, victims of stressful events may

differ quite markedly regarding the valence and the strength of their expressed emotional reactions. Some may cry, some may smile, others may exhibit restlessness or other signs of nervousness, whereas others yet may be neutral and not emit any emotion (Häkkinen-Nyholm, 2017; Magnussen & Wessel, 2010).

To borrow the term used in the clinical psychology setting, the above presented body of research shows us that there is an obvious scientist-practitioner gap between legal psychology research and legal professionals, especially regarding the beliefs legal professionals hold (for presentations on the scientist-practitioner gap as it relates to clinical psychology, see, Cautin, 2011; Lilienfeld et al., 2013; Sobell 2016).

Case Law from the Finnish Supreme Court

The Finnish Supreme Court has given several precedent rulings regarding how judges are to assess the probative value of party and witness statements. To a certain extent the Finnish Supreme Court has referenced legal psychology research and issued guidelines in accordance with this research (references to Christianson & Ehrenkrona, 2011; Finnilä-Tuohimaa, 2009; Granhag & Stridbeck, 2010; Santtila & Weizmann-Henelius, 2008, in KKO 2013:96 and references to Duodecim, 2013; Melinder & Korkman, 2010, in KKO 2013:97; see also KKO 2014:48). Other rulings, however, differ from what we would expect based on legal psychology research. In other words, and in addition to the scientist-practitioner gap presented above, there also exists a scientist-case law gap. Table 1 summarizes the most salient legal guidelines from relevant Finnish Supreme Court rulings pertaining to the probative value of a statement as well as a differentiation of what the individual variables are that affect probative value.

Table 1*Supreme Court Cases, Relevant Legal Guidelines and Probative Value Variables*

Supreme Court case	Legal guideline	Relevant probative value variables
KKO 2013:96	The fact that the injured party's statement is deemed more credible than the defendant's is not enough for a verdict. Even if it seems real and is convincing, it needs indirect evidence to support it. The reliability of witness evidence (fi. <i>henkilötodistelu</i>) cannot be based on the way the person talks, his or her facial expressions or gestures or emotional reactions. What is more important is the statement itself, such as the coherence, realism and constancy in its salient content as well as the amount of detail in it.	Coherence; realism; constancy; detail NOT: The way a person talks; facial expressions; gestures; emotional reactions
KKO 2013:97	The way in which the statement has arisen may have central implications for its probative value. It is important that the right kind of interviewing has been conducted, so that as much verifiable information may be obtained and to ensure that the questions are not leading.	How the statement has arisen; proper interviewing; absence of leading questions
KKO 2017:12	The defendants' statements were incredible and improbable in themselves.	Realism
KKO 2019:54	A statement's consistence, realism, coherence between different times of giving the statement and how detailed the statement is as well as how, in what kind of situation and in what way the memories and the statement regarding the sexual acts have arisen are important components when assessing the probative value of the statement.	Consistence; realism; coherence; different times of giving the statement; how detailed the statement is; what kind of situation and what way the memories and statement have arisen

Table 1 Continued

Supreme Court case	Legal guideline	Relevant probative value variables
KKO 2021:5	The injured party's statement may in itself be enough. The reliability of it must be even more carefully assessed than normal and its faultlessness secured with the help of other, objectively acceptable criteria. Alternative explanations need to be evaluated with due care.	Statement may be reliable; Objectively acceptable criteria; alternative explanations

Live Versus Video-Mediated Statements in Court

Similar to many other countries, a party being heard for probative purposes can be heard using videoconference or other suitable technical means of communication in Finland (Code of Judicial Procedure, Chapter 17, Section 52, in force as of 1 January 2016; see also Ministry of Justice, Finland [OM], 2020). Statements given via videoconference or pre-recorded video versus statements given live in the courtroom have been the subject of some research. From a legal point of view, the most prominent concern raised has been that of whether videoconferencing compromises the parties' constitutional rights (Johnson & Wiggins, 2006; Rowden et al., 2010). Should research show that statements given via videoconference put the statement giver in a worse position than statements given live, videoconferencing would probably be prohibited in the courts altogether.

As regards the medium of the video, there has been some research regarding whether the camera angles or camera shots affect the assessment of witness or party statements and the ensuing sentence (Dahlberg, 2013; Johnson & Wiggins, 2006; Landström, 2008; Rowden et al., 2010; Thielmeyer, 1992); some research regarding the concerns of judges and other legal professionals in relation to videoconferencing (Rowden et al., 2010; Wallace & Rowden, 2018); some research regarding the limitations of communication when hearing someone via videoconference (Wallace & Rowden, 2018); some research regarding how the court's authority is affected when the physical premises and some of the court "rituals" are removed or restricted when hearing someone through videoconference (Dahlberg, 2013; Mulcahy, 2008; Rowden et al., 2010; Wallace & Rowden, 2018;); as well as some research regarding how the social interactions in the courtroom are affected by infusing modern technology into it (Dahlberg, 2013; Wallace & Rowden, 2018; Rowden et al., 2010). While this research and the debate as well as the concerns raised in connection with it are important,

this research has not shown that a statement given via videoconference would put the statement giver in a worse position merely because of the medium of the video.

Some research has also focused on the view of the injured party, for example., how much distress injured parties experience when confronting their attackers or how it feels to recall a traumatic experience, such as rape, in front of the court (Häkkinen-Nyholm, 2017; Kenniston, 2015; Thielmeyer, 1992). This line of research suggests that injured parties experience less distress when heard via videoconference.

Furthermore, there has also been previous research and debate regarding whether actors in the legal system assess statements given via videoconference or pre-recorded video differently from those given live in the courtroom (Johnson & Wiggins, 2006; Kenniston, 2015; Landström, 2012, 2008; Levén & Wersäll, 2011; Perry, 2008; Thielmeyer, 1992). This line of research and debate usually concludes that presentation mode should not matter at all, seeing as there is such strong research evidence showing that deception detection rarely exceeds the level of chance. Landström's (2008) doctoral dissertation contained four studies where truth-telling and lying adult and child witnesses gave statements that were assessed by adult mock jurors with the aim of investigating, among other things, whether there were differences in adults' perception, deception detection, and memory of the statements as a function of presentation mode (live versus video environment) as well as different camera perspectives (close-up, medium shot, long shot, and more). Landström's results indicated that live statements were perceived more positively than statements given via pre-recorded video for both adults and children, and that live statements by children were perceived more positively than those given via videoconference. The results also showed that those who saw a close-up shot were not as ready to believe the child's statement compared to those who saw a long camera shot. In other words, different presentation modes did give rise to different evaluations. As regards deception detection on the other hand, Landström concluded that her studies failed to find an effect where deception detection accuracy would differ between the observers depending on which presentation mode (live / pre-recorded video / videoconference) is used. Taking into consideration the research on deception detection accuracy (or the lack of it) presented above, this is not surprising.

In another study by Landström et al. (2012), Swedish district judges were somewhat more inclined to believe party statements given live than via videoconference or pre-recorded video. Live statements were also assigned slightly more probative value than those given via videoconference or pre-recorded video, supporting the conclusions drawn in Landström,

2008. However, there has also been at least one study where the mode of presentation has not had any impact on assessments made about the statement givers (Ellison & Munro, 2014).

On yet another note, there is also research indicating that deception detection may decrease in accuracy when using visual cues compared to audio or text presentations (Bond & DePaulo, 2006; Dahlberg, 2013). Although deception detection accuracy may decrease, there is research from both psychology and communication sciences showing that visual cues are important in a communicatory sense. The presentation and message may be more vivid and easier to understand when told live, because we also use body language, tone and the likes in communicating our message. They may not be important indicators of deception, but they are still important in communication and making ourselves understood as well as influencing each other (Cordell & Keller, 1993; Dahlberg, 2013; Müller et al, 2013; OM 2020; Rowden et al., 2010). Furthermore, videoconference fatigue and attention issues during videoconference meetings have been shown to be an issue during the recent increase in videoconferences due to the COVID-19 pandemic (Bennet et al., 2021). Technical difficulties may also hamper the hearing, and obviously this is a problem that is not present in live hearings (Powell & Wright, 2009; Rowden & Wallace, 2018).

The Current Study

Terminological Challenges

To study probative value, reliability and credibility in either the legal psychology or the legal context it is important to consider how these terms are used in research and practice. In the Finnish Code of Judicial Procedure (1734), the terms “credible evidence”, “reliable piece of evidence”, “reliability of the piece of evidence”, and “reliable evaluation of the credibility of the person being heard” all appear (Chapter 17 Section 2 subsection 2; Chapter 17 Section 8 paragraph 4; Chapter 17 Section 25 subsection 3; Chapter 17 Section 52 paragraph 3). In the preparatory works, it is stated that it is essential for the court to write down its standpoints regarding the witness’ credibility and the witness statement’s reliability (HE 46/2014 vp).

A distinction where factors regarding the statement giver are assigned to the term “credibility” and factors regarding the statement itself are assigned to the term “reliability” has been quite common in the legal literature (Holmgård, 2019; Landström et al., 2012). However, this distinction has not been used consistently in either scholarly work or Supreme Court practice. Sometimes “credibility” may refer to the person’s willingness to speak the truth whereas “reliability” may refer to the person’s capability of rightly rendering and

expressing what happened (Holmgård 2019). Rasilainen (2006), on the other hand, defines credibility as the correctness of the observation that the person is expressing and doesn't use the term reliability at all. In KKO 2021:5, the term convincing appears alongside the term reliability (par. 98). In KKO 2017:12, the Supreme Court refers to credibility, even though it is clear the "classical" distinction would refer to the matter at hand as one of reliability (see the conclusion that the defendants' statements were unrealistic and not credible, par. 27). Again, in KKO 2013:96, the Supreme Court refers to the credibility of the injured party's statement and not the credibility of the statement giver.

Petterson (2017) concluded that these terminological inconsistencies constitute bad legal concept formation. In her doctoral dissertation, Petterson, now Bergius (2021), suggests not using the term "reliability" and goes forth developing the term "credibility" further, dividing it into three categories: perception, memory and retention. Väisänen and Korkman (2014), to the contrary, suggest not using the term "credibility", seeing as credibility in the Finnish legal context pertains mostly to the nonverbal cues of the statement giver, which, according to legal psychology research, cannot be used to ascertain whether someone is telling the truth or not. Therefore, they suggest using the term "reliability".

Due to these inconsistencies, it was clear to us that we would encounter difficulties in our survey were we to ask judges about credibility and reliability, since the contents of these constructs would vary depending on the judges' definition of the terms. To circumvent this potential source of error, we chose to only use the term "probative value", since irrespectively of how reliability or credibility is defined, the question is ultimately how much probative value judges assign to different variables.

Aims of the Study

The first aim of our study was to increase the above presented body of research from the perspective of Finnish district judges, to see if the predictions that previous research gave rise to held true, both regarding the judges' views on deception detection and whether they assessed live statements differently from those given via videoconference. This was the first study of this kind in a Finnish context. We also conducted statistical analyses to see if the views of the district judges differed as a function of their years in office, seeing as previous research has shown that professionals are more confident (but not more accurate) in their deception detection skills than laypersons, and that this confidence may vary as a function of experience (Porter et al., 2000; Vrij, 2010).

The second aim of our study was to examine what probative value variables district judges considered important, whether there were any differences in the weights given to these different variables and whether there was a difference between how the statement of the injured party vis-à-vis the defendant was assessed. To our best knowledge, there have been no prior empirical studies similar to ours and therefore this second aim of our study was of a more exploratory nature.

Hypotheses

Our hypotheses were the following:

1. We expected district judges to report assigning probative value to parties' body language and emotional expression differently than what research and Supreme Court case law would call for.

Although our first hypothesis could have been extended to several other factors that research has deemed important in the assessment of probative value of party or witness statements, the scope and scientific rigor required for such analysis was far beyond the scope of this thesis and as such, we restricted our hypotheses to the factors we deemed most important for the current study, that is, body language and emotional expression.

2. We expected the district judges with more years in office to report assigning probative value differently to party statements than district judges with fewer years in office, especially regarding body language and emotional expression.

3a. We expected district judges to report assigning probative value differently to party statements given via videoconference than to party statements given live.

3b. We expected the amount of probative value assigned to party statements given via videoconference to vary as a function of how important the judges reported body language and emotional expression to be in the assessment of probative value.

Our third hypothesis followed partly from the assumptions that were behind the first hypothesis, namely, that district judges as a group would not be aware of research in legal psychology regarding deception detection. If the first hypothesis was true, we expected the third hypothesis to be true as well.

4. We expected that there would be a difference in how easy/difficult district judges reported the act of deception detection to be depending on whether it would be done live or via videoconference. We expected this difference to vary as a function of how important the judges generally reported body language and emotional expression to be in the assessment of probative value.

5a. We expected that there would be a difference regarding the importance of the different probative value variables, that is, certain variables would be reported to be more important for the assessment of probative value than others.

5b. We expected that there would be differences in the reported importance of these probative value variables as a function of whether the statement that would be assessed would be given by an injured party or a defendant.

The fifth hypothesis was more exploratory than the previous ones, as we were not aware of any previous empirical research regarding the subject matter.

Method

Ethical Permission

The study received ethical permission by the Åbo Akademi committee for research ethics within the field of psychology. Before accessing the online survey, clients read information about the study and gave their consent to voluntarily participating in the study.

Participants

The sample consisted of 47 district judges from various District Courts in Finland.

Materials

The present study was part of a larger research project done as a research consortium, consisting of Åbo Akademi, the European regional institute in the United Nations Criminal Justice and Crime Prevention programme network HEUNI, Tampere University and the National Courts Administration. Alongside our research questions, the other members of the research consortium investigated different aspects of how common the use of video technology was among Finnish district judges, how the judges reported responding to video-mediated hearings cognitively as well as what kind of training they reported being in need of.

We created an online survey for the present study, containing questions regarding the importance of several variables when evaluating the probative value of party statements, the media environment, deception detection as well as some background questions. These variables are presented in Table 2 and 3. Closed questions on a five-point Likert-scale and a five-point Osgood-scale were used to improve the reliability of the survey and analyzability of the data. The scale and legend used for different survey questions can be seen in Table 3.

Table 2*Judge District, Experience and Number of Hearings via Videoconference*

Variable	<i>n</i>	
Court of Appeal district		
Helsinki	15	
Turku	9	
Eastern Finland	12	
Vaasa	6	
Rovaniemi	5	
Judge experience (years)		
0–5	19	
6–10	9	
11–20	10	
>20	9	
Number of hearings judge has experienced where a party statement was received via VC		
	Defendant's statement via VC	Injured party's statement via VC
	<i>n</i>	<i>n</i>
0	3	2
1–5	10	8
5–10	15	15
11–20	11	13
>20	8	9

Note. VC = videoconference. All District Courts in Finland belong to one of the five Court of Appeal districts listed in the table.

Table 3*Survey Items, Coding, Means and Standard Deviations of Scores*

Variable	Coding	Injured party		Defendant	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Scale A ^a					
Statement contains many details	Details	4.57	0.62	4.36	0.64
Statement is long	Length	2.51	0.93	2.36	0.94
Statement is clear	Clarity	4.09	0.72	4.15	0.66
Statement is realistic	Realistic	4.66	0.48	4.60	0.50
Statement contains parts that are obviously untrue, when it is viewed in light of other evidence	Conflict - Evidence	4.70	0.46	4.77	0.48
Statement is coherent	Coherence	4.43	0.74	4.38	0.71
Statement changes between different phases of the judicial process	Change - Process phases	4.43	0.54	4.45	0.62
Statement changes between different questions	Change - Questions	4.43	0.58	4.38	0.53
Time lapsed between crime and police report	Time since act	3.64	0.94		
How, when and where the statement came to be (e.g., in psychotherapy)	Genesis of story	4.61	0.58	4.24	0.91
Party is very convinced about the correctness of his/her memories	Convinced	2.66	0.98	2.32	0.94
Party is not very convinced	Not convinced	3.53	0.95	3.26	0.99
Party's body language	Body language	2.53	1.27	2.51	1.27
Party's emotional reactions	Emotions	2.28	1.21	2.17	1.13
Scale B ^b					
Less/more probative value to statement given via videoconference compared to live	Live vs. Video (Probative value)	2.98	0.34	2.96	0.30
Easier/more difficult to assess whether party is talking the truth when heard via videoconference compared to live	Live vs. Video (Deception detection)	2.50	0.67	2.47	0.66

Note: Scale A: = Likert scale used for most items in the survey. Scale B = Osgood scale used for only a few items in the survey.

^a 1 = No importance for probative value, 2 = A little importance for probative value 3 = Neither a lot nor a little importance for probative value 4 = Quite a lot of importance for probative value 5 = Very important for probative value.

^b 1 = Videoconference clearly less probative value/more difficult to assess than live, 2 = Videoconference a bit less probative value/more difficult to assess than live, 3 = No difference between live and videoconference, 4 = Videoconference a bit more probative value/easier to assess than live, 5 = Videoconference clearly more probative value/easier to assess than live

Procedure

Prior to the actual data collection, we piloted the survey twice regarding its contents by having two persons that had previously worked as district judges but had later started working with different assignments within the court system fill in the survey in real time and simultaneously give us feedback. We modified the survey based on these feedback discussions. After this, we time-piloted the modified survey thrice to get an estimate of how long it would take to fill out the survey, the result being approximately 30 minutes.

The data was collected via Webropol in April 2021. The National Courts Authority sent links to the online survey to all District Courts, from which the final respondents volunteered to act as participants.

Results

Statistics and Data Analysis

We conducted all statistical analyses using IBM SPSS Statistics for Windows, Version 27.0 (Armonk, NY: IBM Corp.). Prior to analyses, we checked for possible violations of assumptions. Violations are reported below, whereas met assumptions are not reported.

1. Probative value assigned to parties' body language and emotional expression.

The scores on all four of the following variables were not normally distributed, as assessed by Shapiro-Wilk's test ($p < .05$) and visual inspection of Normal Q-Q plots. Due to sample size (Central Limit Theorem) as well as the nature and robustness of the one sample t -test, the analysis was continued.

We conducted one sample t -tests to determine if the sample data differed significantly from the theoretical value of 1, that is, that body language and emotional expression would have no relevance for the assessment of probative value. Four one-sample t -tests were carried out and tested against a Bonferroni-adjusted alpha level of .0125 (.05/4). Results are reported in Table 4.

Table 4*Results of One-Sample t-tests Examining Importance of Body Language and Emotions*

Variable	Injured party					Defendant				
	<i>M</i>	<i>SD</i>	<i>t</i> (46)	<i>p</i>	<i>d</i>	<i>M</i>	<i>SD</i>	<i>t</i> (46)	<i>p</i>	<i>d</i>
Body language	2.53	1.27	8.30	<.001	1.21	2.51	1.27	8.18	<.001	1.19
Emotions	2.28	1.21	7.23	<.001	1.06	2.17	1.13	7.11	<.001	1.04

Note. The *t*-tests compared reported scores to the theoretical value of 1, which represented “no importance for probative value” on the scale used.

2. Association between experience and how probative value is assigned to certain variables. The scores on all four of the following variables were not normally distributed, as assessed by Shapiro-Wilk’s test ($p < .05$) and visual inspection of Normal Q-Q plots. Due to the nature and robustness of the independent-samples *t*-test, the analysis was continued, and the results were interpreted with the limitation of this violation of assumption in mind.

We conducted independent-samples *t*-tests to compare the reports of district judges with more than 6 years in office with district judges with 0-5 years in office on the variables Body language, Emotions, and Live vs. Video (Deception detection). Six independent-samples *t*-tests were carried out and tested against a Bonferroni-adjusted alpha level of .008 (.05/6). Results are reported in Table 5.

Table 5

Results of Independent-Samples t-tests Examining Association Between Experience and Probative Value Assigned

Variable	Injured party						Defendant							
	0–5 years		6+ years		<i>t</i> (45)	<i>p</i>	<i>d</i>	0–5 years		6+ years		<i>t</i> (45)	<i>p</i>	<i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Body language	2.53	1.22	2.54	1.32	-.03	.980	-.01	2.47	1.22	2.54	1.32	-.16	.871	-.01
Emotions	2.16	1.17	2.36	1.25	-.55	.585	-.17	1.95	.97	2.32	1.22	-1.12	.270	-.34
Live vs. Video (Deception detection)	2.63	.62	2.43	.69	.94 ^a	.352	.30	2.56	.62	2.41	.69	.73 ^b	.467	.23

Note. For the Body language and Emotions variables *n*=19 in the 0–5 years group and *n*=28 in the 6+ years group. For the Live vs. Video (Deception detection) variable when the injured party gave their statement via videoconference *n*=16 in the 0–5 years group and *n*=28 in the 6+ years group. For the Live vs. Video (Deception detection) variable when the defendant gave their statement via videoconference *n*=18 in the 0–5 years group and *n*=27 in the 6+ years group. The different group sizes is caused by variation in how many judges had received injured party or defendant statements via videoconference.

^a *t*(42) due to a few judges who never received an injured party’s statement via videoconference.

^b *t*(43) due to a few judges who never received a defendant’s statement via videoconference.

3a. Probative value assigned to statements given live vs. statements given via videoconference. The scores on both of the following variables were not normally distributed, as assessed by Shapiro-Wilk's test ($p < .05$) and visual inspection of Normal Q-Q plots. Due to sample size (Central Limit Theorem) as well as the nature and robustness of the one sample t -test, the analysis was continued.

For Hypothesis 3a, we conducted one sample t -tests to determine if the sample data differed significantly from the theoretical value of 3, that is, that party statements given live are assigned the same amount of probative value as those given via videoconference. Two one sample t -tests were carried out and tested against a Bonferroni-adjusted alpha level of .025 (.05/2). Results are reported in Table 6.

Table 6

Results of One-Sample t -tests Examining Preference for Live or Video

Variable	Injured party					Defendant				
	<i>M</i>	<i>SD</i>	t(43)	<i>p</i>	<i>d</i>	<i>M</i>	<i>SD</i>	t(44)	<i>p</i>	<i>d</i>
Live vs. Video (Probative value)	2.98	.34	-.44	.660	-.07	2.96	.30	-1.00	.323	-.15

Note. The t -tests compared reported scores to the theoretical value of 3, which represented "equal amount of probative value to statement via videoconference compared to live" on the scale used.

Because there was close to zero variability between the different media environments, further analysis in accordance with Hypothesis 3b was not conducted.

4. Connection between perceived difficulty of detecting deception live vs. via videoconference and probative value assigned to body language and emotional expression in general. Multiple linear regressions were calculated to predict how much more difficult/easier the judges reported detecting deception to be live versus via videoconference based on how relevant they reported body language and emotional expression to be in the assessment of probative value in general. Two multiple linear regressions were calculated, with two independent variables each and tested against a Bonferroni-adjusted alpha level of .0125 (.05/4). For both regressions, one leverage value above 0.2 was found. However, as no Cook's Distance values exceeded 1, there was no violation of assumption in this respect.

For detecting deception in the injured party, a significant regression equation was found ($F(2,41) = 5.25, p = .009$), with an R^2_{adj} of .165, which was found to be below Cohen's (1988) convention for a small effect. For detecting deception in the defendant, a significant regression equation was also found ($F(2,41) = 5.26, p = .009$), with an R^2_{adj} of .162, which

was also found to be below Cohen's (1988) convention for a small effect. The coefficients for both models are reported in Table 7.

Table 7

Regression Coefficients of Body Language and Emotions on Difficulty of Deception Detection Live vs. Video

Variable	Model (Injured party)				Model (Defendant)			
	<i>b</i>	<i>SE B</i>	β	<i>p</i>	<i>b</i>	<i>SE B</i>	β	<i>p</i>
Constant	2.92 (2.50, 3.35)	.21		< .001	2.90 (2.46, 3.33)	.22		.001
Body language	-.32 (-.53, -.11)	.10	-.62	.003	-.31 (-.51, -.11)	.10	-.58	.003
Emotions	.17 (-.05, .39)	.11	.31	.123	.16 (-.07, .38)	.11	.26	.171

Note. 95% confidence intervals reported in parentheses. A negative regression coefficient represents a preference for deception detection live when the independent variable is considered important for probative value. A positive regression coefficient represents a preference for deception detection via videoconference when the independent variable is considered important for probative value.

5a. Differences regarding the importance of different probative value variables.

The scores on the probative value variables were not normally distributed, as assessed by the Shapiro-Wilk's test ($p < .05$) and visual inspection of Normal Q-Q plots. Due to sample size (Central Limit Theorem) as well as the nature and robustness of the one-way repeated measures ANOVA, the analysis was continued. There were also a few outliers in the data, as assessed by inspection of the boxplots for values greater than 1.5 box-lengths from the edge of the box. However, there were no extreme data points and no theoretical reason to remove these outliers, and so they were included in the data and the results were interpreted with these outliers in mind. Due to both non-normality and some outliers, the power of the one-way repeated measures ANOVA was decreased but Type I error rate was not substantially affected.

To test whether there were any differences in how the probative value variables were assessed, we conducted one-way repeated measures ANOVAs to compare the effect of probative value variable type on importance for probative value assessment for the injured party and defendant statements. Two one-way repeated measures ANOVAs were conducted with 14 levels and 13 levels respectively on the independent variable probative value variable type and the models were tested against a Bonferroni-adjusted alpha level of .025 (.05/2).

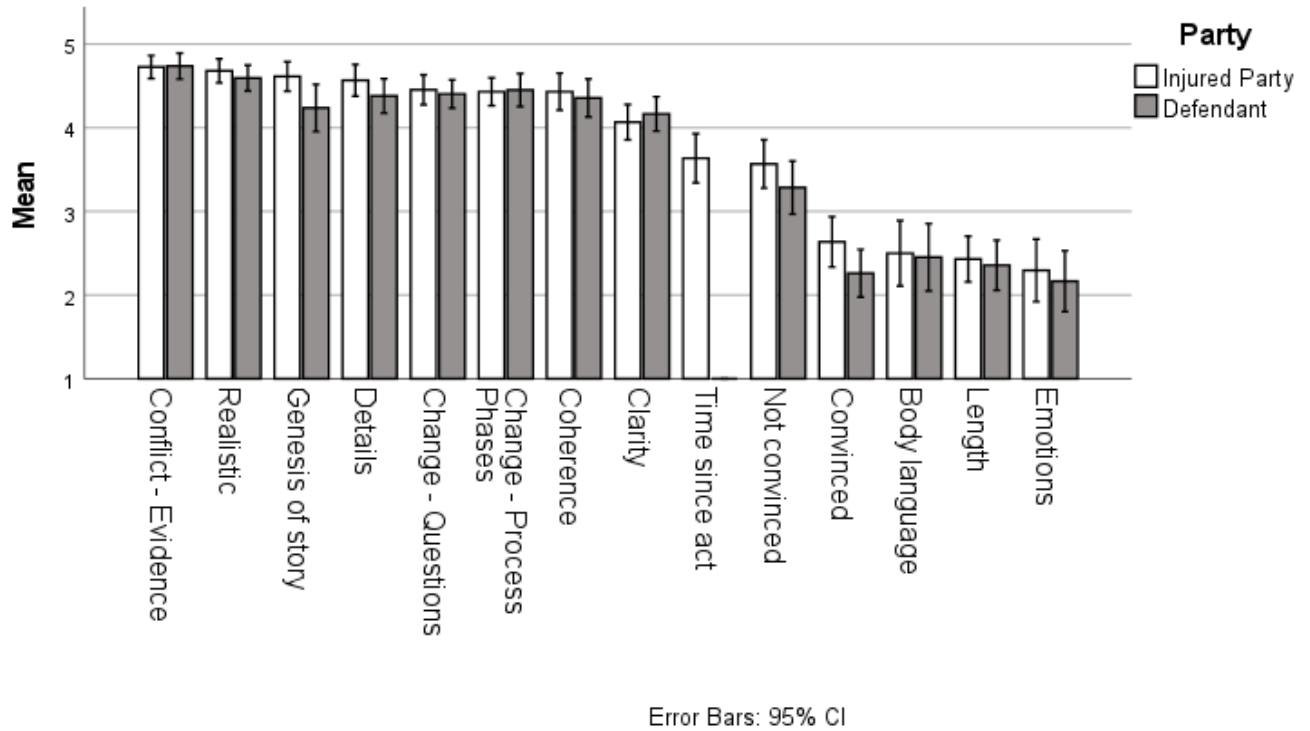
For the one-way repeated measures ANOVA regarding the injured party, Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(90) = 225.29, p < .001$;

therefore, Greenhouse-Geisser corrected tests are reported ($\epsilon = .48$). The results show that the importance for probative value assessment was significantly affected by the probative value variable type, $F(6.26, 269.17) = 63.40, p < .001$.

For the one-way repeated measures ANOVA regarding the defendant, Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(77) = 117.60, p < .001$; therefore, Greenhouse-Geisser corrected tests are reported ($\epsilon = .55$). The results show that importance for probative value assessment was significantly affected by probative value variable type, $F(6.54, 268.43) = 70.93, p < .001$. The levels of the independent variable, that is, the different questions included in the questionnaire, are presented for both injured party and defendant statements in Figure 1, in descending order from most important to least important, ordered according to the scores for the injured party statement.

Figure 1

Comparison of Different Variable Importance for Probative Value



Note. The variable coding used on the horizontal axis is presented in Table 3.

Post hoc analysis with a Bonferroni adjustment revealed several significant pairwise comparisons. The most theoretically relevant of these are discussed in the Discussion chapter.

5b. Differences regarding the importance of different probative value variables as a function of party type (injured party vs. defendant). The scores on all of the following variables were not normally distributed, as assessed by the Shapiro-Wilk's test ($p < .05$) and visual inspection of Normal Q-Q plots. Due to sample size (Central Limit Theorem) as well as the nature and robustness of the paired-samples t -test, the analysis was continued. There were also a few outliers in the data, as assessed by inspection of the boxplots for values greater than 1.5 box-lengths from the edge of the box. However, there were no extreme data points and no theoretical reason to remove these outliers, and so they were included in the data and the results were interpreted with these outliers in mind. Due to both non-normality and some outliers, the power of the paired-samples t -tests were decreased but Type I error rate was not substantially affected.

Hypothesis 5b was tested by comparing the confidence intervals between injured party statements and defendant statements on the one-way repeated measures ANOVAs, and for those confidence intervals that did not seem to overlap or only overlapped to a minor extent (as required for 95% confidence intervals), we followed up by performing paired-samples *t*-tests on these particular probative value variables. Four follow-up paired-samples *t*-tests were carried out and tested against a Bonferroni-adjusted alpha level of 0.0125 (0.05/4)

A study of the confidence intervals in Figure 1 revealed four probative value variable pairs (injured party vs. defendant) that might be significantly different: Convinced, Genesis of story, Details, and Not convinced. Paired-samples *t*-tests were carried out for these four pairs. Results from the paired-samples *t*-tests carried out on these variables are reported in Table 8.

Table 8

Results of Paired-Samples t-tests Comparing Party Statement Probative Value

Variable	Injured party		Defendant		<i>t</i> (46)	<i>p</i>	<i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Convinced	2.66	.98	2.32	.94	4.47	<.001	.65
Genesis of Story	4.62	.58	4.24	.91	3.38 ^a	.002	.52
Details	4.57	.62	4.36	.64	3.53	.001	.51
Not convinced	3.53	.95	3.26	.99	2.66	.011	.39

^a *t*(41) due to respondents having an option to answer "I cannot say" on this survey item.

Discussion

In the current study, we investigated how well certain legal psychology research findings and Finnish Supreme Court rulings regarding the assignment of probative value as well as deception detection in different media environments (live vs. videoconference) have been incorporated in the judicial practice of district judges in Finland. We also investigated whether party statements given for probative purposes are assigned different probative value depending on the media environment in which it is given. Lastly, we also explored whether different variables regarding the party, such as the body language and emotional expression of the party, or variables regarding the party's statement, such as how detailed, how long, or how coherent the party's statement is, are assigned different amounts of probative value by district judges. In connection with this, we also analyzed whether the amount of probative

value assigned to these variables differed for the injured party's and the defendant's statement. To study these questions, we developed an online survey that the National Courts Administration distributed to all District Courts in Finland, and the received survey answers were then analyzed statistically.

The results of the current study suggest a concerning gap between the assessments that district judges do and the scientific evidence from the field of legal psychology regarding how important a party's body language and emotional expression is for the assignment of probative value. As noted above, research has consistently shown that body language and emotional expressions should not have any relevance for the probative value of the statement. These research findings are also reflected in the Finnish Supreme Court ruling stating that body language and emotional expressions should not matter when assessing probative value (KKO 2013:96). Despite this, the respondents of our study clearly deemed these variables to have some importance. Taking the variable body language as example, only around 25% of the district judges in our study gave the "correct" answer as called for by research and case law, that is, that body language does not matter for probative value, whereas around 30% answered that body language either is quite important or very important for the statement's probative value. The remaining 45% answered something in between. On average and contrary to our hypothesis, judges with 0–5 years in office and judges with more than 6 years in office held these same erroneous views to a similar degree.

Interestingly, although the role of body language and emotional expression in the assessment of probative value seems to have been valued far too highly amongst district judges in Finland, they did not report assigning different probative value to statements given live as opposed to statements given via videoconference. As such, Landström et al's (2012) results from Sweden were not replicated in our study. Any other result could have had problematic consequences. Seeing as recent developments in both Finland and Sweden have led or are about to lead to an increased usage of video-recorded witness and party statements in the court process, specifically the stages of appeal, a result showing that presentation mode affects probative value would undermine the foundations of these developments (Dahlberg, 2013; OM 2020; Proposition 2004/05:131). At worst, it would also entail that parties giving their statements in different presentation modes are not being offered legal equality.

Apart from asking how the judges assign probative value to statements given in different media environments, we also asked whether the judges found it easier to assess whether someone is lying live or via videoconference. Responses indicated that judges found

it slightly easier to detect deception live as opposed to via videoconference. Furthermore, our results showed that the more weight judges reported giving to body language when detecting deception, the stronger their preference for live statements was. There is a discrepancy here if this result is compared to the fact that the judges reported assigning equal amount of probative value to statements given live and statements given via videoconference. They reported assigning equal probative value to statements given in the different media environments, but at the same time they reported preferring the live setting when detecting deception. There are at least two explanations for this discrepancy. The first explanation is that judges may separate their mental processes into two dimensions, a factual and a normative dimension. In the factual dimension, the judge would scan for cues of deception through a more intuitive and perhaps automatic process, using such cues as body language, which may be easier to assess in a live setting. In the normative dimension, the result of this scanning would be ignored, and the judge would evaluate the party and their statement through more deliberate reasoning, drawing conclusions and assigning probative value in accordance with the legal framework that the judge considers applicable law. This explanation would form a dual model with some parallels to that of Kahneman's (2011) model of System 1 (fast, automatic, intuitive) and System 2 (slow, deliberate, analytical) thinking. There are two reasons why the first explanation is improbable. First, the questionnaire contained numerous questions regarding probative value and had to be answered from the perspective of the judges acting in their roles as professional judges, not as people casually assessing other people on a more automatic and intuitive basis. Second, the question regarding body language was presented in the context of whether it is important explicitly for the *probative value of the statement*. The second explanation is simply that the judges knew that they should answer that live and video-mediated states are assigned equal probative value but in truth, a slight bias to assess live statements more favorably might still exist.

As regards the relative weights of the different probative value variables, the results show that certain components weigh more than others in the assessment of probative value. These results were not surprising. For example, if parties are subjectively certain that their statements are true, this should obviously not weigh as much as the fact that there is hard evidence speaking against a party statement. The most interesting part of this analysis, however, is two-fold. The fact that we can now present a profile with different probative value variables is valuable in itself, since it dissects the probative value assessment process

into smaller pieces, as judges are required to do (HE 46/2014 vp; KKO 2019:2). This also gives us the possibility to compare the weights given to different variables regarding the statement or statement giver to what research, and case law for that matter, state regarding these variables. Such a comparison reveals that there are certain variables where the alignment of Supreme Court case law, legal psychology research and judicial practice remain incomplete. The most salient of these are the variables of body language and emotional expression, as has been discussed above. Other variables worth highlighting are statement constancy over different utterances as well as coherence. The judges rated these variables to be very important for the probative value of the statement. The Supreme Court of Finland has also ruled that a statement that changes over time should be assessed as having less probative value than a statement which remains the same across process phases and different questions and, likewise, that a statement's coherence is an important variable in the assessment of reliability (KKO 2019:54; KKO 2013:96). According to legal psychological research, the association is not that clear. Truthful and untrue statements may undergo changes in similar amplitudes (Granhag, Strömwall, & Jonsson, 2003). There is also research evidence that the coherence of a statement is not a reliable discriminator between truthful and deceptive accounts (Vrij, 2005). Considering this research, the judges' reported values on these variables (Change - Process phases; Change - Questions; Coherence) are somewhat concerning, as their responses indicate that these variables were quite or very important for the assessment of probative value. The problem, however, is that district judges do not have a choice. The scientist-case law gap, detailed below, prevents them from giving these variables anything but high values.

Judges also tended to value the statements given by injured parties differently from those given by defendants. How convinced parties are over the correctness of their statements, how detailed the statement is, as well as how, when and where the statement came to be (e.g., a statement arising pursuant to psychotherapy), were all considered more important for an injured party statement than a defendant statement. How uncertain parties are over the correctness of their statements was also given somewhat more weight for the injured party compared to the defendant. These results indicate that there is a slight injured party bias when it comes to assessing the probative value of party statements (for a discussion on other, closely related biases in criminal cases, such as confirmation bias and the anchoring effect, see Dahlman, 2018; Imhoff & Nickolaus, 2021; Lidén, Gräns, & Juslin, 2018). Whether this injured party bias can be justified or not quickly turns into a debate in legal

philosophy: In a sense, one might say that the entire accusatory court process in criminal law cases is a form of practical hypothesis testing, where the injured party comes forth with a claim (a hypothesis) that is subsequently tested against alternative hypotheses in the form of alternative courses of events (fi. *vaihtoehtoinen tapahtumainkulku*; Dahlman, 2018; Diesen, 1994; Marjosola et al., 2021; Schelin, 2007). If the injured party's claims "survive" the pre-trial investigation and proceed to deliberation in a main hearing in court, they will, statistically speaking and on average, often have a basis instead of being fabricated arbitrarily (see Dahlman, 2018, who discusses this thought from the perspective of Bayes' theorem; see also Hirvelä, 2006). It remains unclear, however, whether this is enough to grant them a certain precedence over the defense or whether the risk of wrongful conviction when the defendant is not guilty is enough of a counterweight to balance the scales, so that the probative value variables at hand should be given equal weights for both parties. To try to answer these questions would require another thesis but suffice to say that the injured party bias that was revealed by our survey was somewhat surprising and should be the subject of more research, as it may, at worst, endanger party equality.

The Scientist-Case Law Gap

Previous research has shown that there is a scientist–practitioner gap in the context of legal psychology as regards judges' assessments in the context of deception detection (Porter et al., 2000 Vrij, 2010). The present study was in line with this research and showed that this is also the case in Finland. There is also some research indicating that there might be a scientist-practitioner gap in relation to how statements given live or via videoconference are assessed (Landström et al., 2008, 2012). Results from the present study indicate that this gap is not as large as expected, at least in Finland, which is a positive finding for the different actors involved with the judicial process. It is also a good thing considering that in the future, video-recorded statements and videoconferencing will only become more prevalent. The most important finding in our study was that district judges did not report assigning probative value differently to statements given via videoconference compared to those given live, although they did betray a slight preference for live statements when explicitly asked in what media environment they thought deception detection was easier.

Out of the different gaps discussed in this thesis, the most crucial may well be the gap between the science of legal psychology and case law. The reason for this stems from the legal system and its rule that lower-level courts are obliged to follow the rulings of the Supreme Court (for a comprehensive presentation of the Finnish precedent ruling system, see

Launiala, 2020). As such, if the scientist – case law gap is vast, this should lead to a large gap between legal psychology research and practitioners (i.e., district judges) as well. This problem was also observed by Strömwall (2010) in the Swedish legal context, when discussing the fact that the judges of lower-level courts need to apply a method that had been validated by the Swedish Supreme Court but not by legal psychology research. Luckily, both the Swedish and the Finnish Supreme Court have given new rulings since 2010 that are more in line with legal psychology research and have, thus, decreased at least the scientist – case law gap. It still exists, however, and there are certain variables regarding which the Finnish Supreme Court still needs to be made aware of the legal psychology research evidence. Three such variables are constancy of party statements between different process phases (Change – Process phases), constancy of party statements between different questions (Change – Questions) as well as the coherence of the statement (Coherence), as argued above.

Future Research and Practical Implications

At this point, there is already such a substantial body of research evidence on deception detection as well as credibility and reliability assessments that they should guide the probative value assessments in judicial processes. In Finland, many of these theses have already been introduced to the legal profession via Supreme Court cases and by legal scholars. However, there remains a gap between practitioners, legal psychology research and Supreme Court case law, which is concerning. The current study was an attempt to increase awareness of how district judges assess party statements' probative value and how these assessments align with research. This is important, as it can decrease the risk that erroneous beliefs influence the probative value of a statement (Magnussen & Wessel, 2010). Currently, it appears that if a party is giving a statement in court, it would be better to present the kind of body language and the kind of emotions that are congruent with what judges believe is important. If parties do not, they run the risk of giving a statement which is assigned less probative value (see also the meta-analysis by Nitschke et al. (2019) regarding the “emotional victim effect”, where the authors came to a similar conclusion).

While some variables, such as perceived cues to deception, have been the subject of decades of research, other variables still need to be subjected to more study, preferably both experimentally and observationally (see Johnson & Wiggins, 2006, for proposed research designs). After that, we need meta-analyses and reviews that gather all the research evidence so far and present it in a format which is easy to read and understand. This is especially important for legal psychology, because there is also a scientific gap between the discipline

of law and the discipline of psychology (Marjosola, 2021; Stanikić, 2015). Empirical methodology and the scientific-evidential value of different types of research is not widely known amongst lawyers, at least in Finland (Hirvelä, 2006; Marjosola, 2021; see also Schelin, 2006 from a Swedish perspective).

Strömwall (2010) suggests that the Swedish courts should apply the scientific methods of SVA and RM in their assessment of witness or party assessments or, preferably, that the courts would appoint experts to use these methods on the statements at hand. In fact, SVA assessments have been accepted in some courts in North America as well as in parts of Western Europe (Vrij, 2005). However, if one knows how the SVA tool works, it is not difficult to give statements that appear more valid considering the SVA criteria than the statements really are (Vrij, 2005, 2014). As such, the SVA or RM tools in their current forms may not be the answers to our problems regarding probative value assessment of witness and party statements. This has also been noted by Masip et al. (2005), who state that the discriminative power of the RM approach is still too low, as well as by Vrij (2005), who states that the SVA instrument's error rate is still so high that the evidential threshold "beyond reasonable doubt", which is used in criminal cases, cannot be exceeded by using the instrument. That is not to say that the tools are without value, only that we need more research and that the tools need to be cultivated before they can be applied in the court context.

As regards videoconferencing technology, it is likely that it will only continue to improve. Kenniston (2015) states that the videoconferencing tools in Australia are already so developed that you can "see sweat on the forehead of a witness" and predicts that, in the future, there will not be any noteworthy differences between participation live or via video. When this day comes, perhaps the debate regarding the probative value of video-mediated versus live statements will be redundant. To a certain extent it already is. The question whether it is easier or more difficult to detect deception on video versus live is not that relevant, since the only noteworthy difference between these presentation modes is how well one can discern the statement giver, such as his or her body language and emotional expressions. Research clearly shows that no inferences can be made pursuant to these variables anyway.

The reform in the Swedish court system, which started in 2008, and the one currently underway in Finland (OM, 2020) may also be seen as statements from the legislator that video material is not to be assigned any less probative value than material given live, which

was also noted by Dahlberg (2013). Although these reforms do not deal with videoconferencing per se but video-recorded statements from the lower-level courts to be used in later stages of appeal, the take-home message is the same: the medium of the video does not decrease the richness of the statement in such a way that an assessment of its probative value would suffer.

Limitations

There are some limitations that need to be borne in mind when interpreting the results of our study.

As is the case for most empirical sciences, psychological research also quite often suffers from sample selection bias, our study included (see, e.g., Coolican, 2014; Ziliak & Mccloskey, 2008). This mostly stems from the fact that we used a volunteer sample, which may lead to self-selection bias (Sterba & Foster, 2008), which in turn may decrease the generalizability of the results. In our study, this limitation may have been somewhat countered by the fact that our survey was sent through the Finnish National Court Administration, who sent it out to all district judges in Finland.

Some violations of assumptions were found for the statistical analyses. However, these did not cause any direct problems with Type I error rates or with the interpretation of the results, due to either the tests being robust enough to stand the strain of violated assumptions, considering the sample size (Central Limit Theorem), or due to the results all showing the same type of outcome, with highly significant p -levels for all the different tests. For the one-way repeated measures ANOVA, a violation of Mauchly's test of sphericity was corrected for by using Greenhouse-Geisser corrected values. Type I error rates for our statistical analyses were also lowered because we used stringent alpha levels (Bonferroni-corrections for familywise error rate). However, this did entail that the Type II error rates were increased, resulting in low power for the tests used on those of our hypotheses that were subjected to several simultaneous statistical tests. Therefore, the analyses might have failed to find some significant differences between the variables under study as a trade-off of studying numerous variables simultaneously (Field, 2013).

The present study used self-report measures, which sometimes suffer from a decreased accuracy since respondents may not always be aware of their own internal mental processes or because they answer in accordance with their theories of how certain mental processes work. Both context, such as question order, and limitations of memory may also affect survey responses. Furthermore, social desirability bias may distort some self-reported

measures (Holbrook, 2008). However, since the respondents in our survey gave their answers anonymously, social desirability bias should not have been a major issue for our study.

The Likert scale used in our self-report measure may also have suffered from a few limitations. Although widely used and often subjected to statistical analyses in the form of parametric statistical tests, the assumption of equidistance between the Likert points is sometimes not on par with reality and a more conservative approach would be to treat the scores as ordinal data instead of interval level data. The scores on Likert items may also be distorted due to central tendency bias, which may prevent some respondents from selecting the outermost scores on the continuum (Brill, 2008).

Conclusions

The present study investigated certain aspects of the scientist–practitioner gap between legal psychology research and district judges in Finland. District judges reported assigning more probative value to body language and emotional expressions than legal psychology research and Finnish Supreme Court rulings would call for. Regarding certain other variables that are relevant for judges' probative value assessments, our results also highlight a scientist–case law gap. This can be seen in that certain probative value components that the Supreme Court has deemed relevant lack empirical support. Furthermore, our results indicate that there seems to exist both an injured party bias and a slight preference for party statements given live. The preference for party statements given live is somewhat concerning, as it may, at worst, endanger party equality and seeing as the use of video technology in the court room is becoming more common. More research on and refinement of different methods and instruments that may be of help when assessing the probative value of party statements is needed. More effort also needs to be put into making judges and Supreme Courts aware of what current legal psychology research has shown, especially regarding robust scientific results that have been the subject of decades of empirical research.

Swedish Summary – Svensk sammanfattning

Finska tingsdomares bedömningar av partsutlåtanden givna på plats i rätten eller via videokonferens

Forskning och teorier om lögn och lögn-detektion har ökat markant de senaste decennierna, och det publiceras och planeras konstant nya metoder för att skilja mellan sanningsenliga och icke-sanna utsagor (Curtis, 2021; Gamer, 2014; Levine, 2014; McCornack et al., 2014). I dagsläget vet vi dock att endast runt 54 % av bedömningar av huruvida någon talar sanning eller inte är korrekta, ett resultat som knappt är högre än det man skulle få genom att singla slant (Bond & DePaulo, 2006). Olika juridiska aktörer har även blivit alltmer medvetna om forskningsresultatet att lögn-detektion till följd av icke-verbala indikatorer inte är tillförlitligt (HD 2021:5; HD 2019:84; HD 2013:97; HD 2013:96; för en sammanfattning av forskningen inom lögn-detektion, se t.ex. Landström, 2008, 2012; Granhag et al., 2015; Schelin, 2006; Vrij 2014; Väisänen & Korkman, 2014). Trots en ökad medvetenhet bland jurister har forskning ändå visat att olika professionella aktörer inom rättsväsendet ännu kan ha vissa felaktiga övertygelser om lögn-detektion. De har visat sig vara mera självsäkra över sina lögn-detektionsfärdigheter än lekmän, men när dessa färdigheter testats har resultaten varit mer eller mindre på samma nivå som lekmännens resultat (Kassin, Meissner & Norwick, 2005; Landström, 2008; Rosenfeld & Penrod, 2011; Schelin, 2006; Strömwall & Granhag, 2003; Vrij, 2014). Med andra ord finns här en diskrepans mellan forskare och praktiserande jurister (eng. *scientist-practitioner gap*), inte långt olik den som uppmärksamats inom den kliniska psykologin (Cautin, 2011; Lilienfeld et al., 2013; Sobell, 2016).

I Finland är det möjligt att höra en part i bevissyfte med anlitande av videokonferens eller annan lämplig teknisk metod för dataöverföring (Rättegångsbalken 17:52, i kraft sedan 1.1.2016; se även Justitieministeriet [OM], 2020). Sådana utlåtanden givna via videokonferens, samt videoinspelade utlåtanden, har varit föremål för en del forskning. Det största orosmomentet i samband med denna debatt har varit huruvida användningen av videokonferens kan utgöra ett hot mot parternas grundläggande rättigheter (Johnson & Wiggins, 2006; Rowden et al., 2010). Inom detta forskningsområde har man hittat preliminära resultat som tyder på att domare skulle ha en högre beredskap att tro på utlåtanden givna fysiskt på plats än sådana som givits via videokonferens (Landström, 2008; Landström et al., 2012).

Vår studie var ett led i forskningen om partsutlåtandens bevisvärde i olika mediemiljöer samt ett led i forskningen angående vad domare anser är relevant vid bedömningen av bevisvärde och hur detta hänför sig till ovannämnda rättspsykologiska forskningsresultat. Vidare hade vår studie vissa exploratoriska hypoteser.

Våra hypoteser var följande:

1. Tingsdomare förväntades tilldela bevisvärde åt parternas kroppsspråk och känslouttryck annorlunda än vad rättspsykologiska forskningsresultat och Högsta domstolens rättspraxis skulle föranleda.

2. Tingsdomare med mer erfarenhet förväntades tilldela bevisvärde åt partsutlåtanden annorlunda än mindre erfarna tingsdomare, speciellt vad angår kroppsspråk och känslouttryck.

3. Mängden bevisvärde som tilldelas partsutlåtanden förväntades skilja sig mellan utlåtanden givna via videokonferens och utlåtanden givna fysiskt på plats i rättssalen.

4. Det förväntades finnas en skillnad i hur lätt eller svårt tingsdomare anser att lögnedetektion är beroende på huruvida lögnedetektionen sker fysiskt på plats i rättssalen eller via videokonferens. Denna skillnad förväntades ha ett samband med hur viktiga domarna anser kroppsspråk och känslouttryck vara i bevisvärderingen överlag.

5a. Tingsdomare förväntades ge vissa bevisvärdesvariabler mera vikt än andra.

5b. Vikten olika bevisvärdesvariabler får förväntades skilja sig beroende på om utlåtandet ges av målsäganden eller svaranden.

Metod

Etiskt tillstånd

Studien blev godkänd av Åbo Akademis forskningsetiska nämnd för psykologi och logopedi innan datainsamlingen påbörjades. Deltagarna fick information om studien och gav sina medgivande till att frivilligt delta i den.

Deltagare

Samplet bestod av 47 tingsdomare från olika tingsrätter i Finland.

Material och tillvägagångssätt

Studien var en del av ett större forskningsprojekt gjort som ett forskningskonsortium, till vilket Åbo Akademi, HEUNI, Tammerfors universitet samt Domstolsverket hörde.

Vi skapade en online-enkät som innehöll frågor angående vikten av olika variabler när domare tilldelar bevisvärde åt partsutlåtanden, mediemiljön, lögnedetektion och vissa

bakgrundsfrågor. Enkäten piloterades flera gånger, både angående dess innehåll och tiden det tog att fylla i den. Datainsamlingen skedde via Webropol i april 2021. Domstolsverket skickade ut en länk till enkäten till alla tingsrätter i Finland.

Resultat

Vi utförde alla analyser med programvaran IBM SPSS Statistics 27.0. För att testa hypoteserna 1 till 3 utförde vi ett antal t-test. För att testa hypotes 4 utförde vi en multipel regressionsanalys och för att testa hypotes 5 utförde vi en ANOVA med upprepade mätningar. I samband med alla statistiska test granskade vi potentiella brott mot våra antaganden för testen i fråga samt använde oss av Bonferroni-korrigerade alfanivåer. Resultaten rapporteras i tabellerna och texten nedan.

Mängden bevisvärde tilldelat kroppsspråk och känslouttryck

Tabell 1

Resultat av t-test som undersökte hur viktigt parternas kroppsspråk och känslor är

Variabel	Målsägande					Svarande				
	<i>M</i>	<i>SD</i>	<i>t</i> (46)	<i>p</i>	<i>d</i>	<i>M</i>	<i>SD</i>	<i>t</i> (46)	<i>p</i>	<i>d</i>
Kroppsspråk	2,53	1,27	8,30	<,001	1,21	2,51	1,27	8,18	<,001	1,19
Känslor	2,28	1,21	7,23	<,001	1,06	2,17	1,13	7,11	<,001	1,04

Not. t-testen jämförde rapporterade värden med det teoretiska värdet 1, vilket på den använda skalan representerade "ingen betydelse för bevisvärde".

Samband mellan domarerfarenhet och mängden bevisvärde tilldelat vissa variabler**Tabell 2***Result av t-test som undersökte samband mellan erfarenhet och mängden tilldelat bevisvärde*

Variabel	Målsägande					Svarande								
	0–5 år		6+ år		<i>t</i> (45)	<i>p</i>	<i>d</i>	0–5 år		6+ år		<i>t</i> (45)	<i>p</i>	<i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Kroppsspråk	2,53	1,22	2,54	1,32	-,03	,980	-,01	2,47	1,22	2,54	1,32	-,16	,871	-,01
Känslor	2,16	1,17	2,36	1,25	-,55	,585	-,17	1,95	,97	2,32	1,22	-1,12	,270	-,34
Live vs. Video (lögndetek- tion)	2,63	,62	2,43	,69	,94 ^a	,352	,30	2,56	,62	2,41	,69	,73 ^b	,467	,23

Not. För variablerna Kroppsspråk och Känslor var $n=19$ i gruppen 0–5 år och $n=28$ i gruppen 6+ år. För variabeln Live vs. Video (lögndetektion) var gruppstorlekarna liknande, dock fanns det en liten variation och avvikelse för både målsägandeutlåtande och svarandeutlåtande, vilket orsakades av variation i hur många domare som hade tagit emot målsägande- och svarandeutlåtanden via videokonferens.

^a $t(42)$ eftersom några domare aldrig hade tagit emot ett målsägandeutlåtande via videokonferens.

^b $t(43)$ eftersom några domare aldrig hade tagit emot ett svarandeutlåtande via videokonferens.

Skillnad i mängden bevisvärde tilldelat partsutlåtanden givna live eller via videokonferens

Tabell 3

Result av t-test som undersökte preferens för live eller video

Variabel	Målsägande					Svarande				
	<i>M</i>	<i>SD</i>	t(43)	<i>p</i>	<i>d</i>	<i>M</i>	<i>SD</i>	t(44)	<i>p</i>	<i>d</i>
Live vs. Video (bevisvärde)	2,98	,34	-,44	,660	-,07	2,96	,30	-1,00	,323	-,15

Not. t-testen jämförde rapporterade värden med det teoretiska värdet 3, vilket på den använda skalan representerade "lika mängd bevisvärde åt utlåtande givna live som utlåtanden givna via videokonferens".

Skillnad i upplevd svårighetsgrad av att bedöma utlåtanden givna live eller via videokonferens samt samband med mängden bevisvärde tilldelat kroppsspråk och känslouttryck

För lögn-detektion av målsägandens utlåtande hittades en signifikant regressionsekvation ($F(2,41) = 5,252, p = ,009$), med $R^2_{adj} = ,165$, vilket var under Cohens (1988) gränsvärde för en liten effekt. För lögn-detektion av svarandens utlåtande hittades även en signifikant regressionsekvation ($F(2,41) = 5,258, p = ,009$), med $R^2_{adj} = ,162$, vilket även var under Cohens (1988) gränsvärde för en liten effekt. Koefficienterna för båda modellerna rapporteras i tabell 4.

Tabell 4

Regressionskoefficienter för kroppsspråk och känslor på svårighetsgrad att idka lögn-detektion

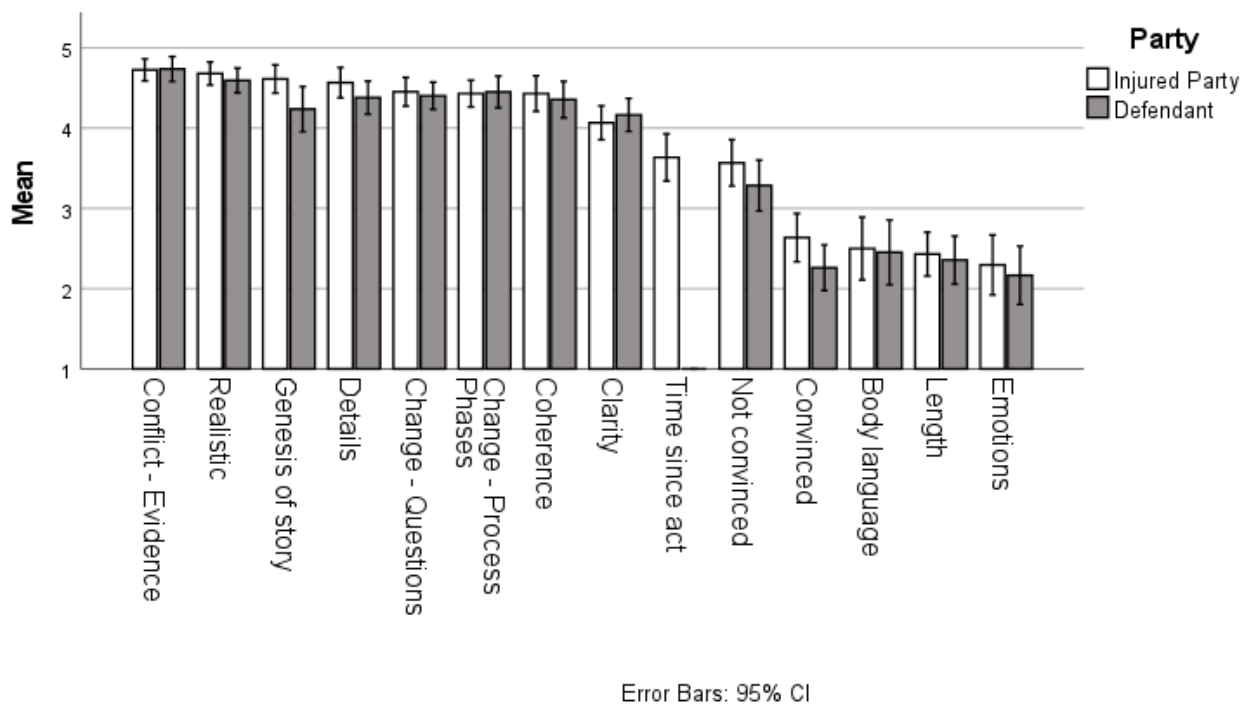
Variabel	Modell (Målsägande)				Modell (Svarande)			
	<i>b</i>	<i>SE B</i>	β	<i>p</i>	<i>b</i>	<i>SE B</i>	β	<i>p</i>
Constant	2,92 (2,50, 3,35)	,21		< ,001	2,90 (2,46, 3,33)	,22		,001
Body language	-,32 (-,53, -,11)	,10	-,62	,003	-,31 (-,51, -,11)	,10	-,58	,003
Emotions	,17 (-,05, ,39)	,11	,31	,123	,16 (-,07, ,38)	,11	,26	,171

Not. 95 % konfidensintervall rapporterade i parenteser. En negativ regressionskoefficient representerar en preferens för lögn-detektion live då när oberoende variabeln anses viktig för bevisvärdet. En positiv regressionskoefficient representerar en preferens för lögn-detektion via videokonferens då när oberoende variabeln anses viktig för bevisvärdet.

Skillnader i mängden bevisvärde tilldelat åt olika bevisvärdesvariabler

Figur 1

Jämförelse av olika variablers betydelse för bevisvärdet



Not. Översättning och förklaring på svenska: Injured Party = målsägande; Defendant = svarande; Conflict - Evidence = motstridighet mellan utlåtande och andra bevis; Realistic = utlåtandets realism; Genesis of story = hur berättelsen i utlåtandet uppkommit (t.ex. via psykoterapi); Details = mängden detaljer i utlåtandet; Change - Questions = förändring i utlåtandet mellan olika förhörfrågor; Change - Process Phases = förändring i utlåtandet mellan olika gånger utlåtandet yttras; Coherence = utlåtandets koherens; Clarity = utlåtandets klarhet; Time since act = tid sedan gärningen; Not Convinced = inte övertygad om egna minnesbildernas exakthet; Convinced = övertygad om egna minnesbildernas exakthet; Body language = partens kroppsspråk; Length = utlåtandets längd; Emotions = partens känslouttryck.

Post hoc-analyser med Bonferroni-korrigerings visade på flera signifikanta parvisa jämförelser. De mest teoretiskt relevanta av dessa diskuteras under rubriken Diskussion.

Skillnader i mängden bevisvärde tilldelat åt olika bevisvärdesvariabler som en funktion av huruvida det är målsägandens eller svarandens utlåtande som är föremål för bedömning

En granskning av konfidensintervallen i Figur 1 visade på fyra par av bevisvärdesvariabler (målsägande vs. svarande) som kunde ha signifikanta skillnader: Övertygelse (Convinced), Hur berättelsen uppkommit (Genesis of story), Detaljer (Details)

och Inte övertygad (Not convinced). Beroende *t*-test utfördes på dessa fyra par. Resultaten rapporteras i tabell 5.

Tabell 5

Result av t-test som jämför typ av partsutlåtande med bevisvärde

Variabel	Målsägande		Svarande		<i>t</i> (46)	<i>p</i>	<i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
Convinced	2,66	,98	2,32	,94	4,47	<,001	,65
Genesis of Story	4,62	,58	4,24	,91	3,38 ^a	,002	,52
Details	4,57	,62	4,36	,64	3,53	,001	,51
Not convinced	3,53	,95	3,26	,99	2,66	,011	,39

Not. Översättning och förklaring på svenska: Convinced = övertygad om egna minnesbildernas exakthet; Genesis of story = hur berättelsen i utlåtandet uppkommit (t.ex. via psykoterapi); Details = mängden detaljer i utlåtandet; Not Convinced = inte övertygad om egna minnesbildernas exakthet.

^a *t*(41) eftersom deltagarna hade möjlighet att svara "Vet ej" på denna enkätfråga.

Diskussion

I vår studie undersökte vi huruvida domare ansåg att mediemiljön i rättssalen (live kontra videokonferens) inverkar på partsutlåtandens bevisvärde samt vilka variabler domare ansåg vara relevanta vid bedömningen av bevisvärde och huruvida de är i linje med rättspsykologiska forskningsresultat. Våra resultat tyder på att det ännu existerar en diskrepans mellan de bedömningar tingsdomare gör av parter kroppsspråk och känslouttryck och hurdana bedömningar de borde göra om man utgår från rättspsykologisk forskning och Högsta domstolens rättspraxis (HD 2013:96). Exempelvis gav endast 25 % av tingsdomarna i vår studie det "korrekta" svaret vad angår betydelsen av kroppsspråk i samband med bevisvärderingen, det vill säga att det inte har någon betydelse. Såväl gruppen domare med 0–5 års domarerfarenhet som gruppen domare med över 6 års domarerfarenhet hade dessa felaktiga uppfattningar. Detta är oroväckande, eftersom det innebär att en part som uppvisar ett sådant kroppsspråk och sådana känslouttryck som är kongruenta med det en domare tror att hör till ett sanningsenligt utlåtande bedöms mera positivt, något som i tidigare forskning kallats "det känslomässiga offret"-effekten (Nitschke et al., 2019).

Intressant nog uppgav domarna att de inte tilldelar bevisvärde annorlunda beroende på om utlåtandet ges fysiskt på plats i rättssalen eller via videokonferens. Vi frågade även huruvida domarna anser att det är lättare att bedöma om en part talar sanning när parten hörs fysiskt på plats i rättssalen jämfört med om parten hörs via videokonferens, och på denna

fråga svarade domarna att de anser det vara något lättare att göra en sådan bedömning live. Denna preferens för live-förhör korrelerade även med hur relevant domarna ansåg kroppsspråk och känslouttryck vara i bevisvärderingen över lag. Det är möjligt att domarna svarat på den första frågan utgående från vad de vet att de borde ha svarat, medan en djupare analys ändå visar på en preferens för live-förhör med resultatet att en part som ger sitt utlåtande via videokonferens i värsta fall kan sättas i en något sämre position i rättsprocessen.

Våra resultat visade vidare att vissa variabler i samband med bevisvärdering bedöms ha mera vikt än andra och vi skapade en profil där dessa viktningar framgår. Nyttan med en sådan profil är att domarnas bevisvärdering kan spjälkas i mindre delar samt att vi kan jämföra dessa viktningar med vad de borde vara i ljuset av rättspsykologisk forskning och Högsta domstolen rättspraxis. Utöver variablerna kroppsspråk och känslouttryck bör variablerna utlåtandets konstanthet mellan olika yttranden samt utlåtandets koherens uppmärksammas. Preliminära rättspsykologiska forskningsresultat antyder att sanningsenliga och icke-sanna utlåtanden går igenom ungefär lika mycket förändring mellan olika gånger då berättelsen yttras (Granhag, Strömwall, & Jonsson, 2003). Det finns även starkt forskningsunderlag som visar att ett utlåtandes koherens inte är en tillförlitlig variabel för att särskilja mellan sanningsenliga och icke-sanna utlåtanden (Vrij, 2005). Tingsdomarna i vårt sampel ansåg ändå att dessa variabler är mycket viktiga för bevisvärderingen. Problemet är dock att tingsdomare inte har något val, eftersom Högsta domstolen i Finland gett riktlinjer enligt vilka dessa variabler är viktiga vid bedömningen av ett utlåtandes tillförlitlighet (HD 2019:54; HD 2013:96), och tingsdomare är bundna av den högsta rättsinstansens prejudikat (Launiala, 2020). Med andra ord finns här utöver en diskrepans mellan forskare och domare (eng. *scientist-practitioner gap*) även en diskrepans mellan forskare och rättspraxis (eng. *scientist-case law gap*).

Slutligen visade våra resultat även att vissa variabler ges mera vikt när de bedöms i samband med målsägandens utlåtande jämfört med svarandens utlåtande. Det är inte lätt att tolka detta resultat, eftersom det även kan vara en konsekvens av hur brottmålsprocesser är uppbyggda (Dahlman, 2018; Diesen, 1994; Marjosola et al., 2021; Schelin, 2007). Hursomhelst är det viktigt att vara medveten om att en sådan här snedvridning i förmån till målsäganden existerar, och det bör vara föremål för framtida forskning, eftersom den i värsta fall kan hota partsjämligheten.

Vår studie hade även vissa begränsningar som bör hållas i åtanke när man tolkar resultaten. Volontärbias kan ha förekommit eftersom vi använde ett sampel av frivilliga

respondenter (Sterba & Forster, 2008) vilket kan påverka generaliserbarheten av resultaten. Vår studie använde även självrapporterade värden vilka kan påverkas av social önskvärdhetsbias och leda till nedsatt exakthet (Holbrook, 2008). Dock gav deltagarna sina svar anonymt, vilket minskar på risken för social önskvärdhetsbias. Vi använde oss av en Likert-skala i samband med enkätfrågorna och antagandet om ekvidistans mellan Likert-poängen motsvarar inte alltid verkligheten. Likert-poäng kan även förvrängas något till följd av central tendensbias (Brill, 2008). Vissa brott mot statistiska antaganden låg för handen, men risken för typ I fel var ändå mycket låg. Detta innebar dock att risken för typ II fel var högre.

Sammanfattningsvis antyder vårt studie att det ännu finns områden gällande domares bevisvärdering där rättspsykologisk forskning, rättspraxis och rättstillämpningen inte står i linje med varandra. Insatser behövs för att göra såväl tingsdomare som Högsta domstolen medvetna om de mest robusta rättspsykologiska forskningsresultaten, så att både objektivitet och partsjämlighet kan tryggas i brottmålsprocesser.

References

- Bennet, A. A., Champion, E. D., Keeler, K. R., & Keener, S. K. (2021). Videoconference fatigue? Exploring changes in fatigue after videoconference meetings during COVID-19. *Journal of Applied Psychology, 106*(3), 330–344.
<https://doi.org/10.1037/apl0000906>
- Bergius, M. (2021). *Karaktärsbevisning - studier av relevansproblematik i personrelaterad bevisning* [Doctoral dissertation, Department of Law]. Lund University.
- Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review, 10*(3), 214–234.
https://doi.org/10.1207/s15327957pspr1003_2
- Brill, J. E. (2008). Likert scale. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods*. SAGE Publications, Inc.
- Cautin, R. L. (2011). Invoking history to teach about the scientist-practitioner gap. *History of Psychology 14*(2), 197–203. <https://doi.org/10.1037/a0022897>
- Christianson, S. Å., Ehrenkrona, M. (2011). *Psykologi och bevisvärdering*. Norstedts Juridik AB.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Coolican, H. (2014). *Research methods and statistics in psychology* (6th ed.). Psychology Press.
- Cordell, L. H., & Keller, F. O. (1993). Pay no attention to the woman behind the bench: Musings of a trial court judge. *Indiana Law Journal, 68*(4), 1199–1207.

https://www.repository.law.indiana.edu/ilj/vol68/iss4/8/?utm_source=www.repository.law.indiana.edu%2Fij%2Fvol68%2Fiss4%2F8&utm_medium=PDF&utm_campaign=PDFCoverPages

Curtis, A. D. (2021) Deception detection and emotion recognition: Investigating F.A.C.E. software. *Psychotherapy Research*, 31(6), 802–816.

<https://doi.org/10.1080/10503307.2020.1836424>

Dahlberg, L. (2013). A modern trial: A study of the use of video-recorded testimonies in the Swedish court of appeal. *Studies in Law, Politics, and Society*, 61, 81–135.

<https://doi.org/10.1108/S1059-4337%282013%290000061007>

Dahlman, C. (2018). *Beviskraft: metod för bevisvärdering i brottmål*. Norstedts Juridik AB.

Diesen, C. (1994). *Bevisprövning i brottmål*. Norstedts Juridik AB

Duodecim. (2013). *Suomalaisen Lääkäriseura Duodecimin ja Suomen*

Lastenpsykiatriyhdistyksen asettaman työryhmän laatima Käypä hoito -suositus lapsen seksuaalisen hyväksikäytön epäilyn tutkimisesta 22.5.2013 [No longer available as of 6 March 2020: <https://www.kaypahoito.fi/lapsen-seksuaalisen-hyvaksikayton-epailyn-tutkimista-koskeva-kaypa-hoito-suositus-poistetaan-julkaisusta-6-3>]. Helsinki.

Ellison, L. & Munro, V. E. (2014). A 'special' delivery? Exploring the impact of screens, live-links and video-recorded evidence on mock juror deliberation in rape trials. *Social & Legal Studies* 2014, 23(1), 3–29. <https://doi.org/10.1177/0964663913496676>

Field, A. (2013). *Discovering statistics using IBM SPSS statistics and sex and drugs and rock'n'roll* [4th ed.]. SAGE Publications Ltd.

Finnilä-Tuohimaa, K. (2009). *Expertise and decision making among clinicians in*

investigations of alleged child sexual abuse [Doctoral dissertation, Department of Psychology]. Turun yliopiston julkaisu, Sarja B, Humaniora.

<http://urn.fi/URN:ISBN:978-951-29-3894-0>

Gamer, M. (2014). Mind reading using neuroimaging: Is this the future of deception detection? *European Psychologist*, 19(3), 172–183. <https://doi.org/10.1027/1016-9040/a000193>

Global Deception Research Team. (2006). A world of lies. *Journal of Cross-Cultural Psychology*, 37(1), 60–74. <https://doi.org/10.1177/0022022105282295>

Granhag, P. A., & Stridbeck, U. (2010). Psychological perspectives on the evaluation of evidence. In P.A. Granhag (Ed.), *Forensic Psychology in Context* (1st ed.). Willan. <https://doi.org/10.4324/9781315094038>

Granhag, P. A., & Strömwall, L. A. (2004). *The detection of deception in forensic contexts*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511490071>

Granhag, P. A., Strömwall, L. A., & Jonsson, A.-C. (2003). Partners in crime: How liars in collusion betray themselves. *Journal of Applied Social Psychology*, 33(4), 848–868. <https://doi.org/10.1111/j.1559-1816.2003.tb01928.x>

Granhag, P. A., Vrij, A., & Verschuere, B. (Eds.) (2015). *Detecting deception: Current challenges and cognitive approaches*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118510001>

HE 46/2014 vp. *Hallituksen esitys eduskunnalle oikeudenkäymiskaaren 17 luvun ja siihen liittyvän todistelua yleisissä tuomioistuimissa koskevan lainsäädännön uudistamiseksi*. <https://www.eduskunta.fi/FI/Vaski/sivut/trip.aspx?triptype=ValtiopaivaAsiat&docid=he+46/2014>

- Hirvelä, P. (2006). *Rikosprosessi lapsiin kohdistuvissa seksuaalirikoksissa*. WSOYpro Oy.
- Holbrook, A. (2008). Self-reported measure. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods*. SAGE Publications, Inc.
- Holmgård, L. (2019). *Bevisning i brottmål*. Norstedts Juridik AB.
- Imhoff, R., & Nickolaus, C. (2021). Combined anchoring: Prosecution and defense claims as sequential anchors in the courtroom. *Legal and Criminological Psychology*, 26(2), 215–227. <https://doi.org/10.1111/lcrp.12192>
- Johnson, M. T. & Wiggins, E. C. (2006) Videoconferencing in criminal proceedings: Legal and empirical issues and directions for research. *Law & Policy* 28(2), 211–227. <https://doi.org/10.1111/j.1467-9930.2006.00224.x>
- Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Kenniston, C. W. (2015). You may now 'call' your next witness: Allowing adult rape victims to testify via two-way video conferencing systems. *Journal of High Technology Law* 16(1), 96–126.
- Landström, S. (2008). *CCTV, live and videotapes: How presentation mode affects the evaluation of witnesses* [Doctoral dissertation, Department of Psychology]. Doctoral Theses from University of Gothenburg. <http://hdl.handle.net/2077/9905>
- Landström, S., Willén, R. M., & Bylander, E. (2012). Rättspraktikers inställning till modern ljud- och bildteknik i rättssalen: En rättspsykologisk studie. *Svensk Juristtidning*, 3, 197–218.
- Launiala, M. (2020). *Korkeimman oikeuden ennakkopäätösjärjestelmä* [eBook]. Alma Talent Oy.

- Levén, S. & Wersäll, F. (2011). En modernare rättegång: Hur har det gått? *Svensk Juristtidning*, 1, 18–31.
- Levine, T. R. (2010). A few transparent liars. In C. T. Salmon (Ed.), *Communication yearbook 34* (1st ed.). Routledge. <https://doi.org/10.4324/9780203846278>
- Levine, T. R. (2014). Truth-Default Theory (TDT): A theory of human deception and deception detection. *Journal of Language and Social Psychology*, 33(4), 378–392. <https://doi.org/10.1177/0261927X14535916>
- Lidén, M., Gräns, M., & Juslin, P. (2018). Self-correction of wrongful convictions: Is there a 'System-level' confirmation bias in the Swedish legal system's appeal procedure for criminal cases?—Part I & II. *Law, Probability and Risk*, 17(4), 311–356. Part I doi: <https://doi.org/10.1093/lpr/mgy018>; Part II doi: <https://doi.org/10.1093/lpr/mgy019>
- Lilienfeld, S. O., Ritschel, L. A., Lynn, S. J., Cautin, R. L., & Latzman, R. D. (2013). Why many clinical psychologists are resistant to evidence-based practice: Root causes and constructive remedies. *Clinical Psychology Review*, 33(7), 883–900. <https://doi.org/10.1016/j.cpr.2012.09.008>
- Magnussen, S., & Wessel, E. (2010). Displayed emotions in court: Effects on credibility judgments. In P.A. Granhag (Ed.), *Forensic Psychology in Context* (1st ed.). Willan. <https://doi.org/10.4324/9781315094038>
- Marjosola, H. (2021). Applying psychological generalisations in testimony evaluation ex officio: Benefits and challenges. *Tidsskrift for Rettsvitenskap*, 134(2–3), 299–332. <https://doi.org/10.18261/issn.1504-3096-2021-02-03-04>
- Marjosola, H., Saranpää, T., & Korkman, J. (2021). Hypoteesit ja vaihtoehdot selitykset rikosasian näytön arvioinnissa. *Lakimies*, 119(3–4), 460–496.

- Masip, J., Sporer, S. L., Garrido, E., & Herrero, C. (2005). The detection of deception with the reality monitoring approach: A review of the empirical evidence. *Psychology, Crime & Law*, 11(1), 99–122. <https://doi.org/10.1080/10683160410001726356>
- McCornack, S. A., Morrison, K., Paik, J. E., Wisner, A. M., & Zhu, X. (2014). Information Manipulation Theory 2: A propositional theory of deceptive discourse production. *Journal of Language and Social Psychology*, 33(4), 348–377. <https://doi.org/10.1177/0261927X14534656>
- Melinder, A., & Korkman, J. (2010). Children's memory and testimony. In P.A. Granhag (Ed.), *Forensic Psychology in Context* (1st ed.). Willan. <https://doi.org/10.4324/9781315094038>
- Mulcahy, L. (2008). The unbearable lightness of being? Shifts towards the virtual trial. *Journal of Law and Society*, 35(4), 464–489. <https://doi.org/10.1111/j.1467-6478.2008.00447.x>
- Müller, C., Cienki, A., Fricke, E., Ladewig, S. H., McNeill, D., & Tessendorf, S. (Eds.). (2013). *Body – Language – Communication: An international handbook on multimodality in human interaction* (Vol. 1). De Gruyter Mouton. <https://doi.org/10.1515/9783110261318>
- Nitschke, F. T., McKimmie, B. M., & Vanman, E. J. (2019) A meta-analysis of the emotional victim effect for female adult rape complainants: Does complainant distress influence credibility? *Psychological Bulletin*, 145(10), 953–979. <https://doi.org/10.1037/bul0000206>
- Perry, H. (2008). Virtually face-to-face: The confrontation clause and the use of two-way video testimony. *Roger Williams University Law Review*, 13(2) 565–594.

https://docs.rwu.edu/rwu_LR/vol13/iss2/9

- Petterson, M. (2017). Trovärdighetens dimensioner. *Tidsskrift for Rettsvitenskap*, 130(1), 3–27. <https://doi.org/10.18261/issn.1504-3096-2017-01-01>
- Porter, S., Woodworth, M., & Birt, A. R. (2000). Truth, lies, and videotape: An investigation of the ability of federal parole officers to detect deception. *Law and Human Behavior*, 24(6), 643–658. <https://doi.org/10.1023/A:1005500219657>
- Powell, M. B., & Wright, R. (2009) Professional's perceptions of electronically recorded interviews with vulnerable witnesses. *Current Issues in Criminal Justice*, 21(2), 205–218. <https://doi.org/10.1080/10345329.2009.12035841>
- Proposition 2004/05:131. *En modernare rättegång - reformering av processen i allmän domstol*. https://www.riksdagen.se/sv/dokument-lagar/dokument/proposition/en-modernare-rattegang---reforming-av-processen_GS03131
- Rasilainen, A. (2006). Muutoksenhakukäsitys ja suullisen todistelun uskottavuus hovioikeudessa. *Lakimies* 104(6), 971–989.
- Rautio, J., & Frände, D. (2020). *Todistelu: oikeudenkäymiskaaren 17 luvun kommentaari* (2nd ed.). Edita Publishing Oy.
- Rosenfeld, B., & Penrod, S. D. (Eds.). (2011). *Research methods in forensic psychology*. John Wiley & Sons, Inc.
- Rowden, E., & Wallace, A. (2018). Remote judging: The impact of video links on the image and role of the judge. *International Journal of Law in Context*, 14(4), 504–524. <http://doi.org/10.1017/S1744552318000216>
- Rowden, E., Wallace, A., & Goodman-Delahunty, J. (2010) Sentencing by videolink in

Australia: Up in the air. *Criminal Law Journal*, 34(6), 363–384.

Santtila, P., & Weizmann-Henelius, G. (Eds.) (2008). *Oikeuspsykologia*. Edita Publishing Oy.

Schelin, L. (2006). *Bevisvärdering av utsagor i brottmål* [Doctoral dissertation, Juridiska institutionen]. Stockholm University.

Schelin, L. (2007). *Bevisvärdering av utsagor i brottmål*. Norstedts Juridik AB.

Sobell, L. C. (2016). Bridging the gap between scientists and practitioners: The challenge before us - Republished article. *Behavior Therapy*, 47(6), 906–919.

<https://doi.org/10.1016/j.beth.2016.11.007>

Sporer, S. L., & Schwandt, B. (2006). Paraverbal indicators of deception: A meta-analytic synthesis. *Applied Cognitive Psychology*, 20(4), 421–446.

<https://doi.org/10.1002/acp.1190>

Stanikić, T. (2015). *Silminnäkijätunnistamisen näyttöarvo* [Doctoral dissertation, Faculty of Law, University of Helsinki]. Suomalaisen Lakimiesyhdistyksen A-sarja.

<http://urn.fi/URN:ISBN:978-951-855-344-4>

Sterba, S. K., & Foster, E. M. (2008). Self-selected sample. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods*. SAGE Publications, Inc.

Strömwall, L. A. (2010). Assessing reliability by analysing the verbal content: the case of Sweden. In P.A. Granhag (Ed.), *Forensic Psychology in Context* (1st ed.). Willan.

<https://doi.org/10.4324/9781315094038>

Strömwall, L. Granhag, P. A. (2003). How to Detect Deception? Arresting the Beliefs of Police Officers, Prosecutors and Judges. in *Psychology, Crime & Law*, 2003, Vol. 9, p.

19-36

Thielmeyer, L. H. (1992). Beyond *Maryland v. Craig*: Can and should adult rape victims be permitted to testify by closed-circuit television? *Indiana Law Journal*, 67(3), 797–852.

<https://www.repository.law.indiana.edu/ilj/vol67/iss3/7/>

Trankell, A. (Ed.). (1982). *Reconstructing the past: The role of psychologists in criminal trials*. P A Norstedt & Söners förlag.

Virolainen, J., & Martikainen, P. (2010) *Tuomion perusteleminen*. Alma Talent Oy.

Vrij, A. (2005). Criteria-Based Content Analysis: A qualitative review of the first 37 studies.

Psychology, Public Policy, and Law, 11(1), 3–41. [https://doi.org/10.1037/1076-](https://doi.org/10.1037/1076-8971.11.1.3)

[8971.11.1.3](https://doi.org/10.1037/1076-8971.11.1.3)

Vrij, A. (2014). 14. Detecting lies and deceit: Pitfalls and opportunities in nonverbal and verbal lie detection. In C. Berger (Ed.), *Interpersonal Communication*. De Gruyter

Mouton. <https://doi.org/10.1515/9783110276794.321>

Vrij, A., Granhag, P.A., & Porter, S. (2010). Pitfalls and opportunities in nonverbal and verbal lie detection. *Psychological Science in the Public Interest*, 11(3), 89–121.

<https://doi.org/10.1177/1529100610390861>

Väisänen, T., & Korkman, J. (2014). Eräitä todistajan kertomuksen arviointiin liittyviä kipukohtia oikeuspsykologisen tiedon valossa. *Defensor Legis*, 5, 721–739.

Willén, R. M., & Strömwall, L. A. (2012). Offenders' lies and truths: An evaluation of the Supreme Court of Sweden's criteria for credibility assessment. *Psychology, Crime &*

Law, 18(8), 745–758. <https://doi.org/10.1080/1068316X.2010.548815>

Zilia, S. T., & McCloskey, D. N. (2008). *The cult of statistical significance: How the*

standard error costs us jobs, justice and lives. University of Michigan Press.

Supreme Court cases

KKO 2013:96. <https://finlex.fi/fi/oikeus/kko/kko/2013/20130096>

KKO 2013:97. <https://finlex.fi/fi/oikeus/kko/kko/2013/20130097>

KKO 2014:48. <https://finlex.fi/fi/oikeus/kko/kko/2014/20140048>

KKO 2017:12. <https://finlex.fi/fi/oikeus/kko/kko/2017/20170012>

KKO 2019:54. <https://korkeinoikeus.fi/fi/index/ennakkopaatokset/1560337301080.html>

KKO 2019:2. <https://korkeinoikeus.fi/fi/index/ennakkopaatokset/1547469194758.html>

KKO 2021:5. <https://korkeinoikeus.fi/fi/index/ennakkopaatokset/k20215.html>