

This is an electronic reprint of the original article. This reprint may differ from the original in pagination and typographic detail.

Finnish district judges' assessments of live versus video-mediated party statements in court
Wilkman, Jonas; Antfolk, Jan; Korkman, Julia

Published in:
Scandinavian Journal of Psychology

DOI:
[10.1111/sjop.13024](https://doi.org/10.1111/sjop.13024)

Published: 01/10/2024

Document Version
Final published version

Document License
CC BY

[Link to publication](#)

Please cite the original version:
Wilkman, J., Antfolk, J., & Korkman, J. (2024). Finnish district judges' assessments of live versus video-mediated party statements in court. *Scandinavian Journal of Psychology*, 65(5), 827-845.
<https://doi.org/10.1111/sjop.13024>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Empirical Article

Finnish district judges' assessments of live versus video-mediated party statements in court

JONAS WILKMAN,¹  JAN ANTFOLEK¹  and JULIA KORKMAN^{1,2} ¹Faculty of Arts, Psychology and Theology, Åbo Akademi University, Turku, Finland²The European Institute for Crime Prevention and Control, affiliated with the United Nations (HEUNI)

Wilkman, J., Antfolk, J. & Korkman, J. (2024). Finnish district judges' assessments of live versus video-mediated party statements in court. *Scandinavian Journal of Psychology*, 65, 827–845.

The increase in remote hearings after the COVID-19 pandemic presents an urgent need to examine how judges assess video-mediated witness and party statements compared with live statements. There is currently a limited body of research on this subject. As for the assessment itself, professionals within the judicial system sometimes believe they can detect deception based on visible cues such as body language and emotional expression. Research has, however, shown that lies cannot be detected based on such cues. The Finnish Supreme Court has also given rulings in accordance with the scientific literature. In this study, we used a survey to investigate how much importance a Finnish sample of district judges ($N = 47$) gave to several variables pertaining to the statement or the statement giver, such as body language and emotional expression. We also investigated the association 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 giving more importance to body language and emotional expression than legal psychology research and Finnish Supreme Court rulings would call for. Our results also indicated that there was a slight bias to assess live statements more favorably than statements given via videoconference, as well as a slight bias in favor of the injured party. More effort must be put into making judges and Supreme Courts aware of findings in legal psychology to avoid biases based on intuitive reasoning where it is contrary to scientific evidence.

Key words: Reliability, credibility, veracity, truthfulness, assessment, deception detection, probative value.

Jonas Wilkman, Faculty of Arts, Psychology and Theology, Åbo Akademi University, Turku, Finland. E-mail: jonas.wilkman@gmail.com

INTRODUCTION

The COVID-19 pandemic led to a sudden increase in the use of remote hearings in courtrooms around the world, and in many countries the practice of online hearings is likely to be much more common post-pandemic than pre-pandemic. This brings with it many changes in how different parties are heard, which can have an impact not only on their experience of courtroom communication but also on how their statements are assessed by the judges (de Vocht, 2022; Legg, 2021; Sanders, 2021; Turner, 2021). The currently existing research on how witness statements are assessed live versus via videoconference shows inconsistent results (reviewed briefly in de Vocht, 2022), and much of the existing literature is outdated. Because recent developments in both Finland and Sweden even before the COVID-19 pandemic have led to increased use of video-recorded witness and party statements in the court process, specifically the stages of appeal, the topic of how presentation mode affects or does not affect the judges' assessment of a statement given in court is of utmost importance (EV 214/2021 vp; HE 133/2021 vp; Dahlberg, 2013; Proposition 2004:131). The primary aims of our study were to assess whether Finnish district court judges assessed live statements differently from those given via videoconference, with a particular focus on deception detection, body language, and emotional expression, as well as to examine how much importance judges gave to several different variables related to both the statement itself and the statement giver. To do this, and to get as ecologically valid data as realistically possible, we developed a survey that was sent out by the Finnish National

Courts Administration to all district judges dealing with criminal cases.

Deception detection among laypersons and legal professionals

There is a large body of research on deception and deception detection from recent decades, and new methods for differentiating between truthful and deceptive accounts are constantly being devised and envisioned (Curtis, 2021; Gamer, 2014; Levine, 2014; McCormack, Morrison, Paik, Wisner & Zhu, 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, that is, barely above chance (Bond & DePaulo, 2006). It has moreover been argued that the approximately 4% that is above chance is due to identifying a small group of liars that are very easy to spot (Levine, 2010, 2014).

Much of the research on deception detection has focused on nonverbal indicators of deception, such as body language or emotional expression, and has clearly shown that no reliable nonverbal indicators of deception exist (for a summary on deception detection research, see, e.g., Landström, 2008; Landström, Willén & Bylander, 2012; Granhag, Vrij & Verschuere, 2015; Schelin, 2006; Strömwall, 2010; Vrij, 2014; Väisänen & Korkman, 2014). 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 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 (Oberlader, Quinten, Banse, Volbert, Schmidt & Schönbrodt, 2021; Trankell, 1982; Volbert & Steller, 2014). SVA has been the focus of much research, and its most salient part is the criteria-based content analysis (CBCA), which is a set of 19 criteria that can be used to assess 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 the CBCA. The lists in the SVA and RM approach contain criteria such as "logical consistency," "superfluous details," "spontaneous corrections," "clarity," "reconstructability," "realism," and so on (see, e.g., Hirvelä, 2006; Masip, Sporer, Garrido & Herrero, 2005; Oberlader, Quinten, Banse, Volbert, Schmidt & Schönbrodt, 2021; Schelin, 2007; Willén & Strömwall, 2012). The models have shown a discriminatory power above chance but have been assessed by some scholars as not being reliable enough to surpass the evidential threshold of "beyond reasonable doubt" (Finnish *ei jää varteenotettavaa epäilyä*; Swedish *utom rimligt tvivel*). Also, many of the studies assessing the discriminatory power of these models, especially older studies, have been criticized for having methodological problems (Masip, Sporer, Garrido & Herrero, 2005; Vrij, 2005).

When these models have been tested, some criteria have received more empirical support than others. In the CBCA, the criterion "quantity of detail" has received substantial support (Vrij, 2005), and a meta-analysis by Oberlader, Quinten, Banse, Volbert, Schmidt, and Schönbrodt (2021) indicates that both the CBCA and RM do distinguish between true and fabricated events. While these findings hold true at a group level, they are not easily used as credibility criteria in individual cases, a problem referred to as the G2i problem (Faigman, Monahan & Slobogin, 2014). Importantly, even the criteria with empirical support may fail to distinguish between truthful accounts and accounts based on false memories (Volbert & Steller, 2014). 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, Sporer, Garrido & Herrero, 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 pretrial investigation contra the main hearing in court, may not be a reliable discriminator between truthful and deceptive accounts. 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). Furthermore, there may be a significant memory loss during the occasionally lengthy time between the pretrial investigation and the court hearing (Westera, Kebbel & Milne, 2013).

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 $d = -0.113$, weighted effect size: Cohen's $d = -0.080$). However, the effect sizes are so small as to be practically irrelevant in the court context, something that was highlighted by the researchers themselves (Sporer & Schwandt, 2006).

Although research strongly shows that there are no reliable cues to deception, people continue to erroneously believe that such cues exist (Bogaard, Meijer, Vrij & Merckelbach, 2016; DePaulo *et al.*, 2003; Luke, 2019). Studies with populations from highly varied backgrounds, including practitioners within the legal field, have shown that the most common cue to deception people look for is gaze aversion (Global Deception Research Team, 2006; Granhag & Strömwall, 2004). One of the reasons that this erroneous belief persists in legal practitioners is likely to be the lack of feedback: Practitioners seldom learn whether their assessments of deception were right or not. Judges, police officers, and other professionals tend to be more confident in their deception detection skills than laypersons, but research has shown that their deception detection is at the same level as that of laypersons (Kassin, Meissner & Norwick, 2005; Landström, 2008; Porter, Woodworth & Birt, 2000; Rosenfeld & Penrod, 2011; Schelin, 2006; Strömwall & Granhag, 2003; Vrij, 2014; Vrij, Granhag & Porter, 2010; see, however, Wright & Wheatcroft, 2017, who argue for a more nuanced perspective regarding police officers).

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). On a more general level, there is a growing body of research regarding what has been termed "the emotional witness effect," which entails that a witness who shows more distress and negative emotional reactions is more readily believed than a witness who is more controlled and neutral in their nonverbal demeanor (Ask & Landström, 2010; Magnussen & Wessel, 2010; Nitschke, McKimmie & Vanman, 2019; van Doorn & Koster, 2019). This is in obvious conflict with the above stated robust research finding that no specific emotional reactions can or should be seen as indicators of truthful or nontruthful accounts.

To summarize the above presented research evidence regarding deception detection: There is robust evidence that there are no reliable nonverbal indicators of deception, and that on average the correctness of assessments of deception is more or less at the level of chance. Some individual criteria regarding the statement itself (and not the nonverbal presentation of the statement giver) have received empirical support in their potential to distinguish between truthful and deceptive accounts. Out of these, "quantity of detail" is the most supported one. Other criteria that traditionally have been thought to be good differentiators, such as "clarity," "coherence," and "statement consistency over time," have not received a lot of empirical support, and to the contrary, there are studies showing that they might not be good

differentiators at all. The more comprehensive statement assessment models of SVA and RM, which include several different criteria, have received some empirical support, but not enough to be eligible for use in courts.

The Finnish legal community has become increasingly aware that detecting deceit is an extremely difficult task, and the Supreme Court of Finland has given several precedent rulings regarding how judges are to assess the reliability and credibility 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. Other rulings, however, differ from what we would expect based on legal psychology research. In Appendix 1, a summary is presented with the most salient legal guidelines from relevant Finnish Supreme Court rulings pertaining to the reliability and credibility of a statement as well as a differentiation of what the Supreme Court deems to be the individual variables that affect a statement's probative value.

Despite an increased awareness of legal psychological issues, lawyers in Finland still do not undergo any systematic, formal training in legal psychology or deception detection. There are some optional studies at the universities, and an optional post-graduation, 2-year specialization education in legal psychology for various professionals working with judicial processes, but no systematic training either in universities or at the courts (University of Helsinki, n.d.; Marjosola, Saranpää & Korkman, 2021; Väisänen & Korkman, 2014). Becoming a district judge in general does not follow a formal or systematic training trajectory in Finland, although most if not all district judges have undergone a 1-year court training as trainee district judges. In 2015, Finland was one of four EU countries without mandatory judge training (HE 270/2016 vp). However, in the preparatory work for the Finnish Courts Act this matter was addressed, and a Judicial Training Board was instated on 1 January 2017 as well as a small percentage of separate educational assistant judge positions (17 positions in 2018), which are filled for 3 years at a time (HE 270/2016 vp; Judicial Training Board, 2018). The purpose of the Judicial Training Board is "to plan and coordinate, jointly with the National Courts Administration and the courts, the training of the staff involved in applying the law at the courts of law, from court traineeships to supplementary training" (Judicial Training Board, 2020). These operations are still in their early stages, though (Judicial Training Board, 2023).

Live versus video-mediated statements in court

Although the COVID-19 pandemic led to an exponential increase in the hearing of witnesses through videoconference or other suitable technical means of communication in court (de Vocht, 2022; Legg, 2021; Sanders, 2021; Turner, 2021), statements given via videoconference or prerecorded video versus statements given live in the courtroom had already been the subject of some research prior to the pandemic (Johnson & Wiggins, 2006; Landström, Ask & Sommar, 2015, 2019; Landström, Willén & Bylander, 2012; Poulin, 2004; Williams, 2011). 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, Wallace & Goodman-Delahunty, 2010).

As regards the medium of the video, several issues have been explored, for example, how videoconferencing technology may limit communication in the courtroom, affect social interactions in the courtroom, or decrease the court's authority as some court "rituals" and physical premises are removed or restricted (Dahlberg, 2013; Mulcahy, 2008; Rowden & Wallace, 2018; Rowden, Wallace & Goodman-Delahunty, 2010). Furthermore, the question of whether actors in the legal system assess statements given via videoconference or prerecorded video differently from those given live in the courtroom has also been investigated (Johnson & Wiggins, 2006; Kenniston, 2015; Landström, 2008; Landström, Ask & Sommar, 2019; Landström, Willén & Bylander, 2012; Levén & Wersäll, 2011; Perry, 2008; Thielmeyer, 1992). Landström's (2008) doctoral dissertation contained four studies in which truth-telling and lying adult and child witnesses gave statements that were assessed by adult mock jurors in different experimental conditions. Her results indicated that live statements were perceived more positively than statements given via prerecorded 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 with 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/prerecorded video/videoconference) was 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, Willén, and Bylander (2012), Swedish district judges were somewhat more inclined to believe party statements given live than via videoconference or prerecorded video. Live statements were also assigned slightly more probative value than those given via videoconference or prerecorded video, supporting the conclusions drawn by Landström (2008). However, there have also been at least a few studies in which the mode of presentation did not have an impact on assessments of the statement givers (Ellison & Munro, 2014; Rossner & Tait, 2023).

On yet another note, there is also research indicating that deception detection may increase in accuracy when using audio-only or text-only presentations, where the one assessing the statement may focus fully on the verbal content of it as opposed to focusing on both verbal content and visual cues (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 presented in an audiovisual manner, 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 for communication and making ourselves understood as well as influencing each other (Cordell &

Keller, 1993; Dahlberg, 2013; Müller, Cienki, Fricke, Ladewig, McNeill & Tessoro, 2013; Rowden, Wallace & Goodman-Delahunty, 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, Campion, Keeler & Keener, 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 deception detection in the legal context, the terms “probative value,” “reliability,” and “credibility” are of relevance, and it is important to consider how these terms are used in research and practice. A distinction in which variables regarding the statement giver are assigned to the term “credibility” and variables regarding the statement itself are assigned to the term “reliability” has been quite common in the legal literature (Holmgård, 2019; Landström, Willén & Bylander, 2012). However, this distinction has not been used consistently. 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 does not use the term “reliability” at all.

Petterson (2017) concludes 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 exclusively 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. When judges engage in deception detection, this ultimately affects the outcome of the case through the legal concept of probative value (Finnish *näyttöarvo*; Swedish *bevisvärde*). The court determines what has been proven and what has not been proven in the case and considers 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; a similar principle is guiding the Swedish judicial procedure [see the Swedish Code of Judicial Procedure, Chapter 35, Section 1, in force as of 1 January 1948]).

“Probative value” is a court’s determination of how strongly the evidence supports that which it is aimed at proving. For example,

how strongly does the injured party’s own statement, presented as evidence, support the alleged fact that the defendant raped the injured party. It is a subjective probability estimate done by the judge: In light of the presented evidence, how high does the judge deem the probability that circumstances have happened in the manner claimed. This is not a mathematical or statistical probability analysis, but a matter of a judge’s subjective certitude. Having said that, the judge still needs to make this estimate thoroughly, impartially, and following objective criteria that justify the end result as well as following “general rules of experience.” The judge’s subjective certitude may not be based on vague intuition or feeling and obviously not on outright arbitrariness (Dahlman, 2018; Frände, Helenius, Hietanen-Kunwald, *et al.*, 2017; KKO 2013:96). When considering the probative value of the presented evidence, 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. In this analysis, the court needs to determine the aggregate probative value of the different pieces of evidence in support of the circumstance that is to be proven and then compare it with the aggregate probative value of the different pieces of counterevidence. Lastly, the court determines whether the presented evidence as a whole yields a level of probative value that surpasses the evidential threshold, which in criminal cases is that the defendant’s guilt needs to be proven “beyond reasonable doubt” (Dahlman, 2018; HE 46/2014 vp; KKO 2019:2; Rautio & Frände, 2020; Virolainen & Martikainen, 2010).

In other words, when district judges assess a party statement’s reliability and credibility, it is always done within the framework of assessing a statement’s probative value. For example, if a judge deems a statement to be unreliable or even false, it will have minor or even nil probative value in proving the alleged crime. If a judge deems a statement to be reliable as well as relevant, the statement’s probative value will be a lot higher and, perhaps, together with other evidence in the case, enough to surpass the evidential threshold for a conviction (as happened in KKO 2013:96). Therefore, to circumvent the potential source of error that using the “classical” terms “reliability” and “credibility” would entail, we chose to use only the term “probative value” in our study.

Aims of the study

The first aim of our study was to assess whether Finnish district court judges assessed live statements differently from those given via videoconference. We were also interested in whether the views of the district judges differed as a function of their years in office, 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 increase with experience (Porter, Woodworth & Birt, 2000; Vrij, Granhag & Porter, 2010).

The second aim of our study was to examine what probative value variables judges considered important, whether there were any differences in the weights given to 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 assessing these questions, and therefore this second aim of our study was of a more exploratory nature.

Hypotheses

We proposed the following hypotheses:

Hypothesis 1: District judges will report assigning probative value to parties' body language and emotional expression differently than what research would call for.

Although our first hypothesis could have been extended to several other factors that research has deemed important in the assessment of party or witness statements, the scope and scientific rigor required for such analysis was far beyond the scope of this study, 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.

Hypothesis 2: The district judges with more years in office will report assigning probative value differently to party statements than district judges with fewer years in office, especially regarding body language and emotional expression.

Hypothesis 3a: District judges will report assigning probative value differently to party statements given via videoconference than to party statements given live.

Hypothesis 3b: The amount of probative value assigned to party statements given via videoconference will 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 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 supported, we expected the third hypothesis to be supported as well.

Hypothesis 4: There will be a difference in how easy/difficult district judges reported the act of deception detection to be depending on whether the assessment of witnesses was conducted live or via videoconference. This difference will vary as a function of how important the judges generally reported body language and emotional expression to be in the assessment of probative value.

Hypothesis 5a: There will be a difference regarding the importance of the different probative value variables, that is, certain variables will be reported to be more important for the assessment of probative value than others.

Hypothesis 5b: There will be differences in the reported importance of these probative value variables as a function of whether the statement to be assessed is 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 from the Åbo Akademi committee for research ethics within the field of psychology. Before accessing the online survey, participants read information about the study and gave their consent to voluntarily take part 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 program network HEUNI, Tampere University, and the National Courts Administration.

We created an online survey for the present study, containing questions regarding the importance of different variables when evaluating the probative value of party statements and presentation mode (live/videoconference) deception detection as well as some background questions. These variables are presented in Tables 1 and 2. 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 2. For the exact wording of the survey questions and answers to the open question, see Appendix 2 and 3, respectively.

Table 1. 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 the judge has experienced in which 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	3	2
1–5	10	10	8
5–10	15	15	15
11–20	11	11	13
>20	8	8	9

Note: All district courts in Finland belong to one of the five Court of Appeal districts listed in the table. *N* = 47. VC = videoconference.

Table 2. Survey items, coding, and 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 elapsed 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 or 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 with live	Live vs. video (probative value)	2.98	0.34	2.96	0.30
Easier/more difficult to assess whether party is telling the truth when heard via videoconference compared with 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.

^a1 = 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.

^b1 = Videoconference clearly of less probative value/more difficult to assess than live; 2 = Videoconference of a bit less probative value/more difficult to assess than live; 3 = No difference between live and videoconference; 4 = Videoconference of a bit more probative value/easier to assess than live; 5 = Videoconference clearly of more probative value/easier to assess than live.

Procedure

Prior to the actual data collection, we piloted the survey twice with former district court judges who filled in the survey in real time and simultaneously provided feedback. We modified the survey based on this feedback. We then 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 min. The data was collected via Webropol in April 2021. The National Courts Administration sent links to the online survey to all district courts in Finland, from which the final respondents volunteered to act as participants. In 2015, there were 510 district judges in Finland (HE 270/2016 vp). A minority of the district judges in Finland do not work with criminal cases, and since our survey pertained only to judges working with criminal cases, our survey's population in 2021 may be estimated to have been around 350 to 400 district judges (personal communication, National Courts Administration of Finland). Out of these, 166 judges opened the survey, 102 of them started answering, and finally 47 of them completed the survey, yielding a response rate of approximately 12–13%.

RESULTS

Statistics and data analysis

We conducted the statistical analyses using IBM SPSS Statistics for Windows, Version 27.0 (Armonk, NY: IBM Corp.). Prior to the 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

We looked at the percentual distribution of answers on the five-point Likert scale. Results are reported in Table 3. The results clearly show that most of the judges in our sample reported assigning probative value to parties' body language and emotional expression differently than what research would call for, which would be that the variables have no importance for probative value, represented by 1 on our Likert scale.

2. Association between experience and how probative value is assigned to certain variables

We conducted independent-samples *t*-tests to compare the reports of district judges with more than 6 years in office with district judges with less than 6 (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 0.008 (0.05/6). Results are reported in Table 4. The results show that we did not find any statistically significant differences in the averages on these variables between judges with 0–5 years in office and judges with more than 6 years in office.

3a. Probative value assigned to statements given live vs. statements given via videoconference

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

Table 3. Percentual distribution of answers regarding the importance of body language and emotions

Variable	Injured party					Defendant				
	1 ^a	2	3	4	5	1	2	3	4	5
Body language (%)	27.7	25.5	17.0	25.5	4.3	27.7	27.7	14.9	25.5	4.3
Emotions (%)	34.0	27.7	19.1	14.9	2.1	36.2	27.7	19.1	12.8	2.1

Note: $N = 47$. Percentual distribution of answers on the five-point Likert scale regarding whether party body language or emotions are seen as important for probative value or not.

^a1 = 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.

Table 4. Results of independent-samples t -tests examining the association between experience and probative value assigned

Variable	Injured party								Defendant							
	0–5 years				6+ years				0–5 years				6+ years			
	M	SD	M	SD	$t(45)$	p	d	M	SD	M	SD	$t(45)$	p	d		
Body language	2.53	1.22	2.54	1.32	−0.03	0.980	−0.01	2.47	1.22	2.54	1.32	−0.16	0.871	−0.01		
Emotions	2.16	1.17	2.36	1.25	−0.55	0.585	−0.17	1.95	0.97	2.32	1.22	−1.12	0.270	−0.34		
Live vs. video (deception detection)	2.63	0.62	2.43	0.69	0.94 ^a	0.352	0.30	2.56	0.62	2.41	0.69	0.73 ^b	0.467	0.23		

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 are 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.

Table 5. Results of one-sample t -tests examining preference for live or video

Variable	Injured party					Defendant				
	M	SD	$t(43)$	p	d	M	SD	$t(44)$	p	d
Live vs. video (probative value)	2.98	0.34	−0.44	0.660	−0.07	2.96	0.30	−1.00	0.323	−0.15

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

videoconference. Two one-sample t -tests were carried out and tested against a Bonferroni-adjusted alpha level of 0.025 (0.05/2). Results are reported in Table 5. The results show that we did not find a statistically significant difference in how much probative value judges assign to statements given live versus statements given via videoconference.

Because there was close to zero variability between the different presentation modes, further analysis in accordance with Hypothesis 3b, that is, the hypothesis that the amount of probative value assigned to party statements given via videoconference would vary as a function of how important the judges reported body language and emotional expression to be in the assessment of probative value, 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 via videoconference versus live (coded as Live vs. Video [Deception detection]) based on how relevant they reported body language and emotional expression (coded as Body language and Emotions, respectively) 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 0.0125 (0.05/4). 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 ($F[2, 41] = 5.25, p = 0.009$), with an R^2_{adj} of 0.165. For detecting deception in the defendant, a significant regression equation was also found ($F[2, 41] = 5.26, p = 0.009$), with an R^2_{adj} of 0.162. The coefficients for

Table 6. 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 to 3.35)	0.21		<0.001	2.90 (2.46 to 3.33)	0.22		0.001
Body language	-0.32 (-0.53 to -0.11)	0.10	-0.62	0.003	-0.31 (-0.51 to -0.11)	0.10	-0.58	0.003
Emotions	0.17 (-0.05 to 0.39)	0.11	0.31	0.123	0.16 (-0.07 to 0.38)	0.11	0.26	0.171

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.

both models are reported in Table 6. These results show that the models were statistically significant and that the models can account for approximately 16% of the variability in how difficult judges perceived detecting deception was via videoconference versus live for the injured party and the defendant, respectively. How much probative value the judges assigned to a party's body language was a statistically significant predictor, whereas how much probative value they assigned to a party's emotional expression was not statistically significant within the model. For every one-point increase on the Likert scale of how important the judges deemed body language to be, the perceived difficulty of detecting deception via videoconference versus live went down 0.32 points (95% confidence interval [CI] -0.53 to -0.11) for the injured party and 0.31 points (95% CI -0.51 to -0.11) for the defendant, representing a preference for deception detection live when body language is deemed important for probative value.

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

There were 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 reasons to remove these outliers, 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 0.025 (0.05/2).

For the one-way repeated measures ANOVA regarding the injured party, Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(90) = 225.29$, $p < 0.001$; therefore, Greenhouse-Geisser corrected tests are reported ($\epsilon = 0.48$). The results show that the importance for probative value assessment was significantly affected by the probative value variable type, $F(6.26, 269.17) = 63.40$, $p < 0.001$.

For the one-way repeated measures ANOVA regarding the defendant, Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(77) = 117.60$, $p < 0.001$; therefore, Greenhouse-Geisser corrected tests are reported ($\epsilon = 0.55$). The results show that importance for probative value assessment was significantly affected by probative value variable type, $F(6.54, 268.43) = 70.93$, $p < 0.001$.

To summarize, the results show that there were significant differences in how important the judges deemed certain variables to be for the assessment of probative value, that is, all variables were not assigned equal importance. 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 Fig. 1, in descending order from most important to least important, ordered according to the scores for the injured party statement.

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

5b. Differences regarding the importance of different probative value variables as a function of party type (injured party vs. defendant)

There were 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 reasons 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 overlapped to only a minor extent (as required for 95% CIs), 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 0.0036 (0.05/14).

A study of the confidence intervals in Fig. 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

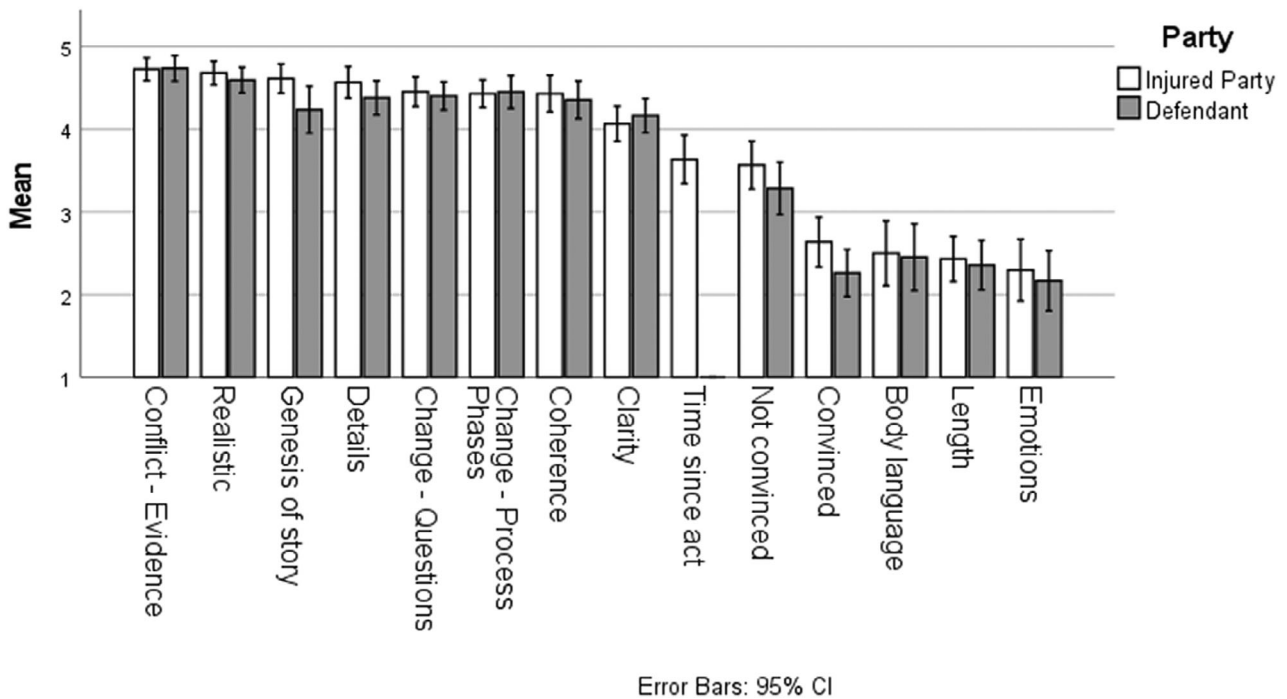


Fig. 1. Comparison of different variable importance for probative value. The variable coding used on the horizontal axis is presented in Table 2.

out for these four pairs. Results from the paired-samples *t*-tests carried out on these variables are reported in Table 7. The results show that there were statistically significant differences in the averages of all these variables except the variable Not convinced, indicating that the variables Convinced, Genesis of Story, and Details are all assigned more probative value when the injured party's statement as opposed to the defendant's statement is being assessed.

DISCUSSION

In the current study, we investigated how well scientific findings (some of which are manifested in certain Finnish Supreme Court rulings) regarding the assessment of party and witness statements in different presentation modes (live vs. videoconference) have been incorporated into the judicial practice of district judges in Finland. Furthermore, we explored whether different variables related to the party, such as their body language and emotional expression, or variables regarding their statement, such as how

detailed, how long, or how coherent their statement is, are assigned different 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 discrepancy between the assessments that district judges do and the scientific evidence from the field of legal psychology regarding how important the judge deems a party's body language and emotional expression to be. As noted above, research has consistently shown that body language and emotional expression should not have any direct relevance when engaging in deception detection or otherwise assessing the reliability and credibility of the statement. These research findings are also reflected, and explicitly referenced, in the Finnish Supreme Court ruling stating that body language and emotional expression should not matter when assessing the probative value of party and witness statements (KKO 2013:96). Despite this, the respondents in our study clearly deemed these variables to have some importance. Less than a third (28%) of the district judges in our study assessed that body language does not matter for probative value, whereas around 30% answered that body language is either quite important or very important for the statement's probative value. The remaining 42% answered something in between. As for the variable emotional expression of the injured party, 34% stated that emotional expression does not matter for probative value, whereas around 17% answered that emotional expression is either quite important or very important for the statement's probative value, with the remaining 49% answering something in between. Furthermore, and contrary to our hypothesis, we did not find

Table 7. 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	0.98	2.32	0.94	4.47	<0.001	0.65
Genesis of story	4.62	0.58	4.24	0.91	3.38 ^a	0.002	0.52
Details	4.57	0.62	4.36	0.64	3.53	0.001	0.51
Not convinced	3.53	0.95	3.26	0.99	2.66	0.011	0.39

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

evidence for differences in the averages on these variables between judges with 0–5 years in office and judges with more than 6 years in office.

Interestingly, although the role of body language and emotional expression in the assessment of probative value seems to have been valued too highly among district judges in Finland, we did not find evidence for a difference in how much probative value they assign to statements given live as opposed to statements given via videoconference. As such, Landström, Willén, and Bylander's (2012) results from Sweden, where Swedish district judges were somewhat more inclined to believe party statements given live than via videoconference, were not directly replicated in our study. Any other result could have had very problematic consequences. Seeing as recent developments in both Finland and Sweden have led 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 (HE 133/2021 vp; Dahlberg, 2013; Proposition 2004:131). 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 judges assign probative value to statements given in different presentation modes, we also asked whether judges found it easier to assess whether someone is giving a deceptive account live or via videoconference. Responses indicated that judges assessed it to be 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 assessing probative value, the stronger their preference for live statements was. This is contradictory to the fact that we did not find evidence for a difference in how much probative value the judges assigned to statements given live versus those given via videoconference. We did not find evidence for a difference in how much probative value they assigned to statements given in the different presentation modes, but at the same time they reported preferring the live setting when detecting deception. Excluding potential methodological reasons for not finding a difference, such as potential issues with the sensitivity of the measure, there are at least two explanations for the contradiction. 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 (faulty) cues as body language, which may be easier to assess in a live setting and thus leading to the answer that it is easier to detect deception live. However, in the normative dimension, the result of this scanning would be ignored, and the judge would evaluate the party and their statement through more slow, deliberate reasoning, drawing conclusions and assigning probative value in accordance with the legal framework that the judge considers applicable law, thus leading to a result in which live and video-mediated statements are assigned equal probative value (see, e.g., Kahneman, 2011, regarding discussions on system 1 and 2 reasoning). There are two reasons why this explanation is improbable. First, the questionnaire contained numerous questions regarding probative value, which had to be answered from the perspective of 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, which seems more likely, is simply that the judges knew that they should not answer that they assign probative value differently to live versus video-mediated statements, but in truth, a slight bias to assess live statements more favorably might still exist, similar to the one found in Landström, Willén, and Bylander's (2012) study.

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, a party who states in court that they are 100% certain that their statement is true should obviously not be given as much weight as the fact that there is hard evidence speaking against their statement, for example, security camera footage clearly and unambiguously showing that the party's statement is not true. The most interesting part of this analysis, however, is twofold. 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 Finnish judges are required to do (HE 46/2014 vp; KKO 2019:2). This gives us the possibility of comparing the weights given to different variables regarding the statement or statement giver with what research, and case law for that matter, state regarding these variables. Such a comparison reveals that there are certain variables for which 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 that 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; Volbert & Steller, 2014). 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 (coded as Change – Process phases, Change – Questions, and Coherence in our material above) 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, as Finnish Supreme Court case law compels them to value these variables highly in their probative value assessments (Marjosola, 2021).

Judges also tended to value the statements given by injured parties differently from those given by defendants. How convinced a party is over the correctness of a statement and 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. 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 that is beyond the scope of this article, but suffice it 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 discrepancy between legal psychology and case law

Out of the different discrepancies discussed in this article, the most crucial may well be the discrepancy 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 discrepancy between legal psychology and case law is large, this should entail that practitioners (i.e., district judges) do not incorporate the findings of legal psychology research in their day-to-day work either. This problem was also observed by Strömwall (2010) in the Swedish legal context, when discussing the fact that 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. Nonetheless, certain variables that are not supported by legal psychology research still remain with full precedential force. Three such variables are constancy of party statements between different process phases (Change – Process phases), constancy of party statements between different questions (Change – Questions), and the coherence of the statement (Coherence), as argued above.

Future research and practical implications

There is already a substantial body of research evidence on the pitfalls of deception detection and assessing witness statements that 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, a concerning discrepancy between legal practitioners, legal psychology research, and Supreme Court case law still remains. The current study was conducted to increase our knowledge 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 judges' assessments of statements (Magnussen & Wessel, 2010). Currently, it appears that judges assign less value to a party statement in court if it is not accompanied by body language and emotions that are congruent with judges' beliefs of how truthful witnesses behave.

As regards videoconferencing technology, it is likely that it will 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 reliability and credibility 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 expression. 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 (HE 133/2021 vp; EV 214/2021 vp) may also be seen as statements from the legislator that video-recorded statements in court are not to be assigned less probative value than those 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; Zilia & 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 National Courts Administration, who sent it out to all district judges in Finland.

The present study used self-report measures, which sometimes suffer from 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, and the most salient questions regarding probative value assessments and deception detection were presented in different and indirect ways, social desirability bias should not have been a major issue for our study. There is also a notion among researchers that judges are a “difficult population to study,” as put by Casaleiro, Relvas, and Dias (2021), who found relatively low response rates in their review of the empirical literature from the last 40 years regarding another theme related to judges and prosecutors. Optimally, a controlled experiment using real judges would be done (instead of the mock jurors that are normally used; see, e.g., Ellison & Munro, 2014; Landström, 2008), but this is unrealistic

considering the difficulty of studying judges just mentioned. As such, the fact that we managed to get 47 real judge participants from a small country such as Finland may be seen as a very good result, even though it remains a relatively small sample size from a methodological point of view.

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

In the present study, we investigated the discrepancy between legal psychological research and district judges' assessments of party statements in Finland. District judges reported giving more importance to body language and emotional expression than legal psychology research and Finnish Supreme Court rulings would call for. Our results also indicate that there is both an injured party bias and a slight preference for party statements given live. This is somewhat concerning since the use of video technology in the courtroom is becoming much more common. More research is needed on different methods and instruments to help assess the probative value of party statements as well as on individual variables that affect probative value and how much importance they should be given in the overall assessment. More effort also must be put into making judges and Supreme Courts aware of findings in legal psychology. Judges need to be able to avoid biases based on intuitive reasoning about what truthful accounts consist of, to the extent they are contrary to scientific evidence.

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>.

CONFLICT OF INTEREST

We have no known conflicts of interest to disclose. This article is based on Jonas Wilkman's master's thesis in psychology.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the 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. Data are available from the authors with the permission of the research consortium.

REFERENCES

- Ask, K. & Landström, S. (2010). Why emotions matter: Expectancy violation and affective response mediate the emotional victim effect. *Law and Human Behavior, 34*(5), 392–401. <https://doi.org/10.1007/s10979-009-9208-6>.
- Bennet, A.A., Campion, 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). *Karaktersbevisning - studier av relevansproblematik i personrelaterad bevisning*. [Doctoral dissertation, Department of Law]. Lund University.
- Bogaard, G., Meijer, E.H., Vrij, A. & Merckelbach, H. (2016). Strong, but wrong: Lay people's and police officers' beliefs about verbal and nonverbal cues to deception. *PLoS One, 11*(6), e0156615. <https://doi.org/10.1371/journal.pone.0156615>.
- 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.
- Casaleiro, P., Relvas, A.P. & Dias, J.P. (2021). A critical review of judicial professionals working conditions' studies. *International Journal For Court Administration, 12*(1), 2. <https://doi.org/10.36745/ijca.334>.
- Coolican, H. (2014). *Research methods and statistics in psychology*, 6th edition. 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%2Ffile%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.
- de Vocht, D.L.F. (2022). Trials by video link after the pandemic: The pros and cons of the expansion of virtual justice. *China-EU Law Journal, 8* (1–2), 33–44. <https://doi.org/10.1007/s12689-022-00095-9>.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K. & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*(1), 74–118.
- 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, 23*(1), 3–29. <https://doi.org/10.1177/0964663913496676>.
- EV 214/2021 vp. Eduskunnan vastaus hallituksen esitykseen suullisen todistelun tallentamista ja vastaanottamista yleisissä tuomioistuimissa koskevan lainsäädännön uudistamiseksi https://www.eduskunta.fi/FI/vaski/EduskunnanVastaus/Sivut/EV_214+2021.aspx.
- Faigman, D.L., Monahan, J. & Slobogin, C. (2014). Group to individual (g2i) inference in scientific expert testimony. *The University of Chicago Law Review, 81*(2), 417–480. <https://doi.org/10.2139/ssrn.2298909>.
- Frände, D., Helenius, D., Hietanen-Kunwald, P., Hupli, T., Koulu, R., Lappalainen, J. et al. (2017). *Prosessioikeus*, 5th edition. Talentum Media Oy.
- 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. & 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 133/2021 vp. Hallituksen esitys eduskunnalle suullisen todistelun tallentamista ja vastaanottamista yleisissä tuomioistuimissa koskevan lainsäädännön uudistamiseksi https://www.eduskunta.fi/FI/vaski/HallituksenEsitys/Sivut/HE_133+2021.aspx.
- HE 270/2016 vp. Hallituksen esitys eduskunnalle laeiksi tuomioistuinlain ja eräiden muiden lakien muuttamisesta https://www.eduskunta.fi/FI/vaski/HallituksenEsitys/Sivut/HE_270+2016.aspx.
- 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>.
- Judicial Training Board. (2018). Tuomarinkoulutuslautakunnan toimintakertomus vuodelta 2018 <https://oikeus.fi/tuomioistuimet/fi/index/tuomioistuinlaitos/lautakunnat/tuomarinkoulutuslautakunta/toimintakertomukset.html>.
- Judicial Training Board. (2020). Judicial Training Board <https://oikeus.fi/tuomioistuimet/en/index/tuomioistuinlaitos/lautakunnat/tuomarinkoulutuslautakunta.html>.
- Judicial Training Board. (2023). Tuomioistuinten henkilöstön osaamisen kehittämisen strategia 2019–2024 <https://oikeus.fi/tuomioistuimet/fi/index/tuomioistuinlaitos/lautakunnat/tuomarinkoulutuslautakunta/koulutuksenkehittaminen/osaamisenkehittamisenstrategia2019-2024.html>.
- Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Kassin, S.M., Meissner, C.A. & Norwick, R.J. (2005). “I’d know a false confession if I saw one”: A comparative study of college students and police investigators. *Law and Human Behavior*, 29(2), 211–227.
- 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., Ask, K. & Sommar, C. (2015). The emotional male victim: Effects of presentation mode on judged credibility. *Scandinavian Journal of Psychology*, 56, 99–104. <https://doi.org/10.1111/sjop.12176>.
- Landström, S., Ask, K. & Sommar, C. (2019). Credibility judgments in context: Effects of emotional expression, presentation mode, and statement consistency. *Psychology, Crime & Law*, 25(3), 279–297. <https://doi.org/10.1080/1068316X.2018.1519828>.
- Landström, S., Willén, R.M. & Bylander, E. (2012). Rättspraktikers inställning till modern ljud- och bildteknik i rättsalen: En rättspsykologisk studie. *Svensk Juristtidning*, 3, 197–218.
- Launiala, M. (2020). *Korkeimman oikeuden ennakkopäätösjärjestelmä [eBook]*. Alma Talent Oy.
- Legg, M. (2021). The COVID-19 pandemic, the courts and online hearings: Maintaining open justice, procedural fairness and impartiality. *Federal Law Review* 2021, 49(2), 161–184. <https://doi.org/10.1177/0067205X21993139>.
- 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. <https://doi.org/10.1093/lpr/mgy019>.
- Luke, T.J. (2019). Lessons from pinocchio: Cues to deception may be highly exaggerated. *Perspectives on Psychological Science* 2019, 14(4), 646–671. <https://doi.org/10.1177/1745691619838258>.
- 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). Todistajankertomuksen luotettavuus: Henkilötodistelun selvittäminen, vastaanottaminen ja arviointi oikeuspsykologisen tutkimuksen valossa [Doctoral dissertation, Faculty of Law]. Doctoral Theses from University of Helsinki <http://hdl.handle.net/10138/334339>.
- Marjosola, H., Saranpää, T. & Korkman, J. (2021). Hypoteesit ja vaihtoehtoiset selitykset rikosasian näytön arvioinnissa. *Lakimies*, 119, 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>.
- McCormack, 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>.
- 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. (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>.
- Oberlader, V.A., Quinten, L., Banse, R., Volbert, R., Schmidt, A.F. & Schönbrodt, F.D. (2021). Validity of content-based techniques for credibility assessment: How telling is an extended meta-analysis taking research bias into account? *Applied Cognitive Psychology*, 35(2), 393–410. <https://doi.org/10.1002/acp.3776>.
- 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.
- Pettersson, M. (2017). Trovärdighetens dimensioner. *Tidsskrift för Rettsvetenskap*, 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>.
- Poulin, A. (2004). Criminal justice and videoconferencing technology: The remote defendant. *Tulane Law Review*, 78(4), 1089–1168.

- 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) En modernare rättegång – reformering av processen i allmän domstol. 131. https://www.riksdagen.se/sv/dokument-lagar/dokument/proposition/en-modernare-rattengang---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 edition. Edita Publishing Oy.
- Rosenfeld, B. & Penrod, S.D. (2011). *Research methods in forensic psychology*. John Wiley & Sons, Inc.
- Rossner, M. & Tait, D. (2023). Presence and participation in a virtual court. *Criminology & Criminal Justice*, 23(1), 135–157. <https://doi.org/10.1177/17488958211017372>.
- 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. <https://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.
- Sanders, A. (2021). Video-hearings in Europe before, during and after the COVID-19 pandemic. *International Journal For Court Administration*, 12(2), 3. <https://doi.org/10.36745/ijca.379>.
- 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.
- 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>.
- 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. & Granhag, P.A. (2003). How to detect deception? Arresting the beliefs of police officers, prosecutors and judges. *Psychology, Crime & Law*, 9, 19–36.
- 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>.
- 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.
- Turner, J.I. (2021). Remote criminal justice. *Texas Tech Law Review*, 53(2), 197–272. <https://doi.org/10.2139/ssrn.3699045>.
- University of Helsinki. (n.d.). *Oikeusnotaarin koulutusohjelma: Temaattiset kokonaisuudet*. University of Helsinki, Faculty of Law. <https://www.helsinki.fi/fi/koulutusohjelmat/oikeusnotaarin-koulutusohjelma/opiskelu/temaattiset-kokonaisuudet>.
- Väisänen, T. & Korkman, J. (2014). Eräitä todistajan kertomuksen arviointiin liittyviä kipukohtia oikeuspsykologisen tiedon valossa. *Defensor Legis*, 5, 721–739.
- van Doorn, J. & Koster, N.N. (2019). Emotional victims and the impact on credibility: A systematic review. *Aggression and Violent Behavior*, 47, 74–89. <https://doi.org/10.1016/j.avb.2019.03.007>.
- Virolainen, J. & Martikainen, P. (2010). *Tuomion perusteleminen*. Alma Talent Oy.
- Volbert, R. & Steller, M. (2014). Is this testimony truthful, fabricated, or based on false memory? Credibility assessment 25 years after Steller and Köhnken (1989). *European Psychologist*, 19(3), 207–220. <https://doi.org/10.1027/1016-9040/a000200>.
- 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/10.1037/1076-8971.11.1.3>.
- 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>.
- Westera, N.J., Kebbel, M.R. & Milne, B. (2013). Losing two thirds of the story: A comparison of the video-recorded police interview and live evidence of rape complainants. *Criminal Law Review*, 4, 290–308. <https://doi.org/10.1080/1068316X.2010.548815>.
- 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>.
- Williams, R.A. (2011). Videoconferencing: Not foreign language to international courts. *Oklahoma Journal of Law and Technology*, 7(1), 3. <https://digitalcommons.law.ou.edu/okjolt/vol7/iss1/3>.
- Wright, C. & Wheatcroft, J.M. (2017). Police officers' beliefs about, and use of, cues to deception. *Journal of Investigative Psychology and Offender Profiling*, 14(3), 307–319. <https://doi.org/10.1002/jip.1478>.
- 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.

Received 14 August 2023, accepted 5 April 2024

APPENDIX 1

Table A1. 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 (Finnish <i>henkilötodistelu</i>) cannot be based on the way the person talks or on 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 The way a person talks, his or her facial expressions or gestures, and his or her emotional reactions are not relevant probative value variables
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; in what kind of situation and in what way the memories and statement have arisen
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 in itself; objectively acceptable criteria; alternative explanations

APPENDIX 2

Table B1. Full survey items

Survey item	Coding in Table 2	Answer options
I generally assign the following circumstances importance when assessing the probative value of the injured party's statement (irrespective of whether the party is being heard live in court or via videoconference):		
1. The injured party's statement contains many details	1. Details	<ul style="list-style-type: none"> • No importance for probative value • A little importance for probative value • Neither a lot nor a little importance for probative value • Quite a lot of importance for probative value • Very important for probative value
2. The injured party's statement is long	2. Length	
3. The injured party's statement is clear	3. Clarity	
4. The injured party's statement is realistic	4. Realistic	
5. The injured party's statement contains parts that are obviously untrue, when it is viewed in light of other evidence	5. Conflict – Evidence	
6. The injured party's statement is coherent	6. Coherence	
7. The injured party's statement changes between different phases of the judicial process (e.g., pretrial investigation – main hearing)	7. Change – Process phases	
8. The injured party's statement changes between different questions (e.g., main examination – cross-examination)	8. Change – Questions	
9. The injured party himself/herself is very convinced that their memories are correct	9. Convinced	
10. The injured party himself/herself is not very convinced that their memories are correct	10. Not convinced	
11. The injured party's statement regarding the alleged crime has been presented after a lot of time has passed since the time of the act (e.g., because reporting the alleged crime to the police was delayed)	11. Time since act	
I generally assign the following circumstances importance when assessing the probative value of the defendant's statement (irrespective of whether the party is being heard live in court or via videoconference):		
12. The defendant's statement contains many details	12. Details	<ul style="list-style-type: none"> • No importance for probative value • A little importance for probative value
13. The defendant's statement is long	13. Length	

(continued)

Table B1. (continued)

Survey item	Coding in Table 2	Answer options
14. The defendant's statement is clear	14. Clarity	<ul style="list-style-type: none"> • Neither a lot nor a little importance for probative value
15. The defendant's statement is realistic	15. Realistic	
16. The defendant's statement contains parts that are obviously untrue, when it is viewed in light of other evidence	16. Conflict – Evidence	<ul style="list-style-type: none"> • Quite a lot of importance for probative value • Very important for probative value
17. The defendant's statement is coherent	17. Coherence	
18. The defendant's statement changes between different phases of the judicial process (e.g., pretrial investigation – main hearing)	18. Change – Process phases	<ul style="list-style-type: none"> • Quite a lot of importance for probative value • Very important for probative value
19. The defendant's statement changes between different questions (e.g., main examination – cross-examination)	19. Change – Questions	
20. The defendant himself/herself is very convinced that their memories are correct	20. Convinced	<ul style="list-style-type: none"> • Quite a lot of importance for probative value • Very important for probative value
21. The defendant himself/herself is not very convinced that their memories are correct	21. Not convinced	
Item formulation	Coding in Table 2	Answer options
Questions with separate answers for the injured party and the defendant:		
22. The way in which the memories and the statement came to be is an important factor when you assess a statement given by a party for probative purposes (e.g., did the memories regarding what happened and the subsequent police report arise in psychotherapy or in connection with discussions with different people, or did they arise in discussions or interrogations containing leading questions).	22. Genesis of story	<ul style="list-style-type: none"> • No importance for probative value • A little importance for probative value • Neither a lot nor a little importance for probative value • Quite a lot of importance for probative value • Very important for probative value
23. The body language (gestures, facial expressions, tone of voice, etc.) of the party being heard for probative purposes is an important factor when assessing the probative value of the statement.	23. Body language	<ul style="list-style-type: none"> • No importance for probative value • A little importance for probative value • Neither a lot nor a little importance for probative value • Quite a lot of importance for probative value • Very important for probative value
24. The emotional expression (e.g., crying, getting angry, nervousity, etc.) of the party being heard for probative purposes is an important factor when assessing the probative value of the statement.	24. Emotions	<ul style="list-style-type: none"> • No importance for probative value • A little importance for probative value • Neither a lot nor a little importance for probative value • Quite a lot of importance for probative value • Very important for probative value
25. In general, my experience is that I assign less/more probative value to a party's statement given for probative purposes via videoconference as compared with a statement given live in court.	25. Live vs. video (probative value)	<ul style="list-style-type: none"> • Clearly less probative value • A bit less probative value • Neither less nor more probative value • A bit more probative value • Clearly more probative value
26. In general, my experience is that it is more difficult/easier to assess whether the party being heard for probative purposes is speaking the truth when heard via videoconference as compared with being heard live in court.	26. Live vs. video (deception detection)	<ul style="list-style-type: none"> • Clearly more difficult to assess • A bit more difficult to assess • Neither more difficult nor easier to assess • A bit easier to assess • Clearly easier to assess

APPENDIX 3

Table C1. Answers to open survey question: "Anything else important when assessing the probative value of the injured party's/defendant's statement, what?"

Respondent Nr	Answer regarding injured party (IP)	Answer regarding defendant (D)
Respondent 1	"Is the chain of events presented in the statement realistic in comparison with other evidence in the case."	"Is the defendant's description of what happened possible and probable enough in light of the other evidence."
Respondent 2	"Whether they give their own story or focus on arguing why the opposite party's story is untrue."	"Whether they give their own story or focus on arguing why the opposite party's story is untrue."
Respondent 6	"Does other evidence, direct or indirect, support the statement, for example, has the injured party spoken about the events after they happened to other nonpartisan persons or organizations."	"Assessing the reliability of the defendant's statement is governed by the same fundamental principles as the assessment of statements given by others being heard."
Respondent 7	"As a starting point the same rules and fundamental principles should apply when assessing the defendant's and injured party's statements. Their positions in the judicial process and potential interests as parties to the case may explain why they talk about matters in a certain way. The injured party's statement may beget different meaning in case, either in the injured party's statement or in the other court material, there can be found, e.g., strong antipathy towards the defendant, or information that there have been disagreements between the parties in the past. On the other hand, if there has been a settlement between the parties, this may lead, among other things, to the fact that some circumstances will be emphasized whereas other will be minimized (e.g., in cases of domestic violence)."	"As a starting point the same rules and fundamental principles should apply when assessing the defendant's and injured party's statements. Their positions in the judicial process and potential interests as parties to the case may explain why they talk about matters in a certain way."
Respondent 8	"Matters that arise in free narrative, especially details, are more important than those that arise only if specifically asked of them."	"Whether they give their own story or focus on arguing why the opposite party's story is untrue."
Respondent 10	"One needs to assess whether the statement feels truthful, is coherent and convincing. People behave differently in different situations, and usually one cannot draw any conclusions based on the behavior of the injured party during the alleged crime or in court."	"The presumption of innocence. For that, it is enough that the statement is possible considering other evidence in the case. It doesn't have to be credible; it is enough that it isn't incredible."
Respondent 14	"Can the statement be influenced by, for example, fear, desire for revenge, or by other people? How good is the person at expressing matters (age, language difficulties, high degree of nervousity). As regards the previous question regarding emotions [Question coded as Emotions in Table 2], I do not think other things make a difference for probative value, but nervousity may be a thing to consider at times."	"The same as I wrote for the injured party regarding nervousity."
Respondent 16	"It should overall be required that the injured party separate between observations and other matters that the injured party has not observed but has an opinion of and therefore tells them as true. Actual observations can be compared with other evidence. The comparison may support or weaken the injured party's statement. The observations should specifically be related to the discrepancy between the injured party's and the defendant's statements. One may also pay attention to whether the statement giver takes time to consider before telling detailed circumstances, and, as such, circumstances that can be verified from the injured party's perspective. If the party takes time to consider, for example by presenting counter questions, this may indicate that there is an attempt to control, either partly or in whole, the consistency between a false statement and its different versions. There is likely not a need to do this regarding a truthful experience. It seems that laying forth a false statement is harder if the one giving the statement is in the same physical premises as a counterparty who knows the truth, but this conflicts with the need to protect the injured party. Particularly in relation to sex crimes, I have met several injured parties who have put forth false allegations and even falsified technical evidence, and then admitted to it when becoming exposed. There is reason to examine the potential motives of the injured party as well as how the story arose. The motives for a potential false allegation cannot	"In principle, hearing the defendant is governed by similar considerations as hearing the injured party. However, especially in very serious crimes, the defendant regularly has a strong interest in denying their guilt regardless of the veracity of the charges brought against them and without having any specific motivation regarding the case. In addition, in the pretrial investigations there have been so many different kinds of procedures in view of the defendant's right to due process that at times one has had to regard for example the defendant's changing of their statement a bit more mildly."

(continued)

Table C1. (continued)

Respondent Nr	Answer regarding injured party (IP)	Answer regarding defendant (D)
	necessarily be guessed or even explainable in the eyes of outsiders but might be 'internal' matters of the injured party. The allegation may not originally have been intended to reach the authorities but may have arisen through a different motive. It is hard to back out of an allegation after it has been brought forth, even if one wants to. Regarding potential emotional reactions, the judge needs to understand the difference between an observation and a conclusion. A tear in the eye is an observation, but the veracity of the statement is a conclusion drawn from that tear and not an observation. Even in the Supreme Court's practice, there may be observed a contradiction between the officially acclaimed principles and their application regarding this. For example, in connection with some of the aforementioned false statements, the injured party has (in my opinion) shown real distress and crying, but the emotional reactions were not due to the crime. Due to fear of being exposed or there being an innocent person under allegation, the situation may be very scary and anxiety provoking to the injured party, also in case they are being heard to investigate the veracity of a false allegation."	
Respondent 18	"The specificity of individual memories in relation to other memories."	"The specificity of individual memories in relation to other memories."
Respondent 22	"The statement gets a bit more value if the injured party also describes circumstances that hurt their case."	–
Respondent 23	"In light of current knowledge, abundance of detail does not necessarily increase the reliability of the statement."	–
Respondent 27	"Is there a motive to lie?"	"The right to refrain from investigating one's own guilt"
Respondent 28	–	"Defendants seem to be telling more 'freely' about events when being heard via videoconference. They seem to sometimes forget that they have already told about events in the pretrial investigation and that the alternative course of events becomes less probable if their statement in court is completely different from what it was during the pretrial investigation. Sometimes the defendant's lawyer that is physically present in the courtroom becomes surprised when the client wants to be heard via videoconference after all, and then starts to tell stories."
Respondent 31	"How does the statement relate/adapt to undisputed facts in the case?"	"How does the statement relate to undisputed facts in the case?"
Respondent 33	"The relationship between the defendant and injured party and how well they know each other is an important factor when assessing the credibility of the statement."	"The relationship to co-defendants and injured parties may influence the statement."
Respondent 34	"How the statement relates to other indirect evidence, KKO 2013:96, paragraph 7, as well as how the injured party's statement relates to the defendant's statement."	–
Respondent 36	"Does the injured party also bring forth circumstances that hurt their case or do they admit that they don't remember something. In other words, the objectivity of the statement."	"Does the defendant also bring forth circumstances that hurt their case and benefit the injured party's case or do they admit that they don't remember something. In other words, the objectivity of the statement."
Respondent 41	"Especially when it comes to compensation for suffering, the emotional expression is of relevance when assessing how the act has made the injured party's life 'more difficult.'"	"Some of the defendants genuinely show regret, but that is when they have admitted their guilt. I also take notice when a defendant is sitting 'expressionless' vs. an aggressive way of speaking or if a defendant shows how the act was done (e.g., the movement of the hand when striking with a knife), in which case I focus on how the use of force looks especially if/when the use of force has been aggressive."
Respondent 42	"Eye contact and the quality and style of the questions as well as the questioner's expressions, tone of voice, reactions, style and how they treat the one being heard in every phase of the procedure."	"Eye contact and the quality and style of the questions as well as the questioner's expressions, tone of voice, reactions, style and how they treat the one being heard in every phase of the procedure."
Respondent 44	"What kind of impression does one get from the statement as a whole, is it genuine, avoidant, etc. Why is the injured party's statement avoidant?"	"Why is the defendant's statement avoidant/undetailed?"

(continued)

Table C1. (continued)

Respondent Nr	Answer regarding injured party (IP)	Answer regarding defendant (D)
Respondent 45	“How the statement came to be during the pretrial investigation if someone appeals to it after the statement has changed in the pretrial investigation. Has the interrogation been open and comprehensive, and the statement is of course always weighed in relation to other evidence.”	“If the defendant has kept silent during the pretrial investigation, one needs to consider the relation between this silence and the alternative course of events that has been presented in court.”
Respondent 46	“– that which has been told freely, without ‘milking’”	“– that which has been told freely, also circumstances that hurt their case”
Respondent 47	“Need to clarify whether the statement is based on their own observations, drawn conclusions, or what they have heard from others. This is something that is not focused on nearly enough during the pretrial investigation. Quite often the pretrial investigation statement does not enable any conclusions to be made as to what extent the statement is based on own observations, drawn conclusions, and to what extent on what they have heard from others.”	“Same considerations that apply to the injured party. Quite often the defendant’s statement seems to be memorized. In other words, before the main hearing in court, the defendant has probably quite precisely acquainted themselves with what it was they said during the pretrial investigation.”