

ARTICLE

An Empirical Analysis of Credibility Assessment in German Asylum Cases

Dr. Björnstjern Baade*  and Leah Gölz*

Freie Universität Berlin, Berlin, Germany

*Corresponding authors: b.baade@fu-berlin.de, leahgoelz@web.de

(Received 18 May 2022; accepted 13 September 2022; first published online 13 March 2023)

Abstract

This study analyzes empirically how 236 German court decisions assess the credibility of asylum seekers' accounts of their persecution. In their reasoning, the courts rely on generally accepted content-based credibility criteria, including consistency, level of detail, and timeliness of the claim. But they also rely on conduct-based criteria, which have been resoundingly discredited in the relevant scientific literature. Too rarely, the courts considered confounding factors such as cultural distance or interpreter mistakes. They need to be more aware of their duty to confront applicants with negative credibility criteria. Article 4 (5) Qualification Directive played no role whatsoever in the sample analyzed in this study, which can be explained by specifics of German asylum law.

The human judgment that is required in the balancing of credibility criteria and confounding factors is problematic for its subjectivity but unavoidable. Attempts at replacing this human credibility assessment with seemingly objective technical means have led to arbitrary decisions and encroached gravely on applicants' human rights. While the credibility assessment procedure employed in German courts is far from flawless, it can produce convincing decisions. It should be further refined and provided with safeguards to arrive at decisions that are as rational and objective as possible.

Keywords: Refugee; Asylum; Credibility; Qualification Directive; Migration; Quantitative

A. Introduction

A person's recognition as a refugee under the 1951 Geneva Refugee Convention¹ frequently depends on the question whether the applicant's account of their persecution is credible.² Decision-makers need to decide whether the account given by the applicants can form the factual basis on which to assess the risk of persecution, should they be returned to the country of origin.

¹Convention Relating to the Status of Refugees, July 28, 1951, 189 U.N.T.S. 137; Protocol Relating to the Status of Refugees, Jan. 31, 1967, 606 U.N.T.S. 267.

²Note on terminology: The asylum procedure in Germany is conducted first before the German Federal Agency for Refugees and Migration and then reviewed by administrative courts. Technically, the person who is an "applicant" in the administrative procedure before the Federal Agency becomes a "claimant" in the court proceedings. The applicant is heard by the Agency and again by the courts in "hearings," which are legally different institutions ("Anhörung" and "informativische Anhörung" respectively). But they serve the same purpose which will be detailed below. While the person taking the decision at the Agency is a decision-maker ("Entscheider"), in court proceedings a judge decides. In order to focus on the credibility assessment, which does in principle not differ between Agency and courts, a uniform terminology will be used: The applicant is heard in asylum interviews. A decisionmaker will assess the account's credibility. This shall be taken to refer to both, the Agency as well as the courts, unless otherwise indicated.

To determine whether the applicant's account in an asylum case is credible, i.e., whether it should be believed for the purposes of the proceedings, so-called credibility criteria are used; in particular, the consistency of the applicant's account and its plausibility. While difficult to apply under the best of circumstances, these criteria are often subject to distortive factors. For example, mistakes by the interpreter or in the transcript of the applicant's interview can produce the false impression of inconsistencies. Credibility criteria and distortive factors can be ambivalent, contradictory, and thus need to be balanced in the individual case to arrive at a decision. Like other legal balancing processes, credibility assessment is criticized for its subjectivity (B.). Credibility criteria are well-established in the case law of many jurisdictions, and not only in asylum law. But research on the issue of credibility determination has mostly been qualitative so far (C.).

This article empirically analyzes a sample of German asylum cases that is available in the most authoritative German legal database Juris.³ Such an empirical approach is particularly called-for in this area because the number of decisions that are published is so large that it cannot be comprehensively analyzed in a qualitative fashion. In fact, there are so many cases that even an empirical approach must be limited in scope. While the database does not contain all decisions that are actually taken, Juris is the most relevant legal database that is used by courts and practitioners, shaping their view of the law. The approach followed here can shine a light on a larger part of that case law and thus produce insights that might otherwise be missed.

This article is the result of a quantitative assessment of 236 German asylum cases that dealt with credibility assessment in the year 2017 (D.). The reasoning which the German administrative courts gave in these cases was analyzed, *first*, to determine which credibility criteria were actually used in practice. *Second*, it was assessed whether and to what extent confounding factors that might distort the application of credibility criteria were taken into account by the judge according to the decision's reasons. Concluding Observations on the limits of objectivity in decision-makers will close this article (E.).

B. Credibility Assessment in German Asylum Cases

This Section will introduce the credibility criteria (also called "reality criteria")⁴ that are generally accepted in the literature as well as in the practice of various national and international courts (I.). The confounding factors that could potentially distort their assessment will be described thereafter (II.). The Section will close by pointing out the balancing character of this assessment and the purpose that asylum interviews serve in this regard (III.).

I. Credibility Criteria

The credibility criteria that are used to determine whether the applicant's account should be believed for the purposes of refugee status determination can be divided into two general categories: Those criteria that are based on the account's content (1.) and those based on the applicant's conduct, during the interview or otherwise (2.).

1. Content-Based Criteria

The credibility assessment conducted by German administrative courts is based on a premise that is known in Germany as the "*Undeutsch* hypothesis." Developed by and named after the psychologist *Udo Undeutsch* (1917–2013),⁵ this hypothesis states that there are differences *in content*

³JURIS, www.juris.de (last visited Apr. 14, 2022).

⁴Cf. Gregor Noll, *Credibility, Reliability, and Evidential Assessment*, in OXFORD HANDBOOK OF REFUGEE LAW 609, 614 (Cathryn Costello, Michelle Foster, Jane McAdam eds., 2021).

⁵See also Barbara Hausmann, *Aussagepsychologie und Glaubhaftigkeitsbeurteilung*, in RECHTSPSYCHOLOGIE 93, 95 (Denis Köhler ed., 2014); ALDERT VRIJ, DETECTING LIES AND DECEIT 227 (2nd ed. 2008). For a brief overview of the history of the psychology of witnesses that began before *Undeutsch*, see generally Julia Shaw, Lisa Öhman & Peter van Koppen, *Psychology and Law: The Past, Present, and Future of the Discipline*, 19 PSYCH., CRIME & L. 643 (2013).

between a factual account that is given by a person who *did* actually experience what that person describes, and an account by a person who describes a situation which that person did *not* experience.⁶ Simply put, the hypothesis is based on the assumption that it is difficult to lie in a manner that does not recognizably differ from a subjectively true account with regard to its content.⁷ Something that was really experienced can, in principle, be retold from memory. Invented accounts must be, at least partially, constructed from one's general knowledge about the world.⁸ This is, for example, why seemingly unimportant details are mostly avoided when lying.⁹ Outside of Germany, empirical research starts from a similar premise, namely that it is, in general, more emotionally stressful and/or more cognitively challenging to lie than to tell the truth.¹⁰

Regularly referencing a standard work authored by practicing judges, i.e. *Bender et al.*, “Tatsachenfeststellung vor Gericht” (*Fact-finding Before Courts*),¹¹ German administrative courts consider that the “*Undeutsch* hypothesis” and the credibility criteria based on it have “proven themselves.”¹² It is accepted as the basis of all credibility determinations conducted by German courts, either explicitly as a “scientifically confirmed insight” or at least implicitly by the application of content-related credibility criteria—which this study provides clear confirmation for.¹³ The German Federal Agency for Refugees and Migration (*Bundesamt für Migration und Flüchtlinge*—in the following: Federal Agency) trains its decision-makers on this basis.¹⁴ Outside of Germany, criteria similar to those expounded in *Bender et al.* are generally accepted.¹⁵

But despite their practical importance, the validity of the *Undeutsch* hypothesis and its credibility criteria has, in fact, not been entirely confirmed in empirical research. In some controlled experiments, credibility assessments following these criteria were wrong about 30 % of the time.¹⁶ However, such studies conducted the credibility assessment based on a written interview transcript, which makes the application of credibility criteria considerably more difficult. Such a study design omits the arguably most important aspect of credibility assessment: The asylum interview which serves an indispensable function in this regard.¹⁷ But at least a general connection between a truthful account and positive credibility criteria has been confirmed.¹⁸ The most evidence seems to currently exist for the criterion relating to an account's level of detail.¹⁹

⁶GABRIELE JANSEN, ZEUGE UND AUSSAGEPSYCHOLOGIE, paras. 16, 30 (3rd ed. 2022).

⁷ROLF BENDER, ROBERT HÄCKER & VOLKER SCHWARZ, TATSACHENFESTSTELLUNG VOR GERICHT, paras. 325, 354 (5th ed. 2021).

⁸*Id.* para. 323.

⁹*Cf.* Bundesgerichtshof [BGH] [Federal Court of Justice] July 30, 1999, ENTSCHIEDUNGEN DES BUNDESGERICHTSHOFS IN STRAFSACHEN [BGHSt] 45, 164–182, para. 20.

¹⁰See VRIJ, *supra* note 5, at 43–55, 103–104.

¹¹See generally BENDER, HÄCKER & SCHWARZ, *supra* note 7.

¹²See generally Bundesweite Arbeitsgemeinschaft der psychosozialen Zentren für Flüchtlinge und Folteropfer [BAFF], BEGUTACHTUNG TRAUMATISIERTER FLÜCHTLINGE 58 (2nd ed., 2014) (translation by the authors). See also Verwaltungsgerichtshof Baden-Württemberg [VGH] [Higher Administrative Court of Baden-Wuerttemberg], Case No. A 11 S 2526/17, para. 7, (Nov. 22, 2017); Verwaltungsgericht Berlin [VG] [Administrative Trial Court of Berlin], Case No. 8 K 794.16 A, para. 22 (Jan. 23, 2019); Verwaltungsgericht Gießen [VG] [Administrative Trial Court of Gießen], Case No. 1 K 2449/11.GI.A, para. 29, (Feb. 25, 2014); Hessischer Verwaltungsgerichtshof [VGH] [Higher Administrative Court of Hesse], Case No. 12 UE 2936/93, para. 46, (Aug. 8, 1994).

¹³See VG Berlin, Case No. 8 K 794.16 A para. 22; VG Gießen, Case No. 1 K 2449/11.GI.A para. 29; *cf.* Verwaltungsgericht München [VG] [Administrative Trial Court of Munich], Case No. M 2 K 13.30275, para. 30, (Nov. 29 2013). See also BAFF, *supra* note 12, at 54.

¹⁴See Stephanie Schneider, *Becoming a Decision-Maker*, in ASYLUM DETERMINATION IN EUROPE 295, para. 11 (Nick Gill & Anthony Good eds., 2019) (based on observation of the training of Federal Agency decision-makers).

¹⁵This is shown by the “CREDO” project, jointly conducted by the International Association of Refugee Law Judges (IARLJ), the EU Asylum Support Office (EASO), the UN High Commissioner for Refugees (UNHCR), the Hungarian Helsinki Committee, and further NGOs and academics. See IARLJ, ASSESSMENT OF CREDIBILITY 32 (2013).

¹⁶See VRIJ, *supra* note 5, at 252–523.

¹⁷See *infra* Part II.3.

¹⁸See VRIJ, *supra* note 5, at 223–227, 256–527.

¹⁹See VRIJ, *supra* note 5, at 228 f.; HANS-UDO BENDER, MERKMALSKOMBINATIONEN IN AUSSAGEN 166 (1987).

All of the content-based credibility criteria that are applied in the asylum systems of Germany and many other states are an emanation of the following two general criteria: Consistency and plausibility. *First*, decision-makers take into consideration whether an account is *consistent*, i.e. free from logical contradiction: With itself, *internally*, but also with other evidence, *externally*. For example, if an attack is described, in a first interview before the Federal Agency as having been conducted with “long sticks” and then, before the reviewing court, as conducted with “long knives,” this is, *prima facie*, a logical contradiction.²⁰ *Second*, it is relevant whether an account is *plausible*. In its most general form, plausibility assesses the likelihood of certain events or circumstances, in particular, how coherent an account seems. For example, “the sheer improbability of one individual wrestling himself from a guard, leaving his clothes in the guard’s hand, then evading another five of them, vaulting a two-metre wall, with no one shooting at him, even to wound him, or shouting for others to come” is a factor that speaks against the account’s credibility.²¹

While plausibility in its most general form is often heavily criticized as a credibility criterion for its apparent subjectivity, three more specific emanations of the plausibility criterion are generally accepted and play an important role in the practice of refugee status determination: The *level of detail* and the *level of knowledge* that can be expected from the applicant, as well as the *timeliness of the claim*.

If someone really experienced a certain situation, they will, in general, be able to supplement the account with further information. This information concerns, *first*, *details of the specific event* or the general circumstances that a person claims to have experienced. But this can also concern, *second*, more general, *context-independent factual knowledge*. For example, if someone states that they handled sensitive political and financial issues for a political party, they should be able to describe their activities with more than just platitudes and could be expected to have some general knowledge about finances.²² Details and knowledge that can be expected from a credible applicant are particularly important in cases concerning persecution due to one’s sexual orientation²³ or religious conversion²⁴. In these often difficult cases, the person’s very identity or inner conviction are at issue. For the former, a model was developed by the attorney S. *Chelvan*,²⁵ which is accepted by the UN High Commissioner for Refugees (UNHCR) as well as the European Asylum Support Office (EASO).²⁶ It structures personal experiences that non-heterosexual persons regularly make into certain elements that can be taken into account in credibility assessment: Difference, stigma, shame, and harm (DSSH model).²⁷

Third, an “untimely” claim is known in German asylum law as an “increased” claim (*Steigerung*). This means that the account that was initially given is “increased” either by adding

²⁰Verwaltungsgericht Münster [VG] [Administrative Trial Court of Muenster], Case No. 9 K 2580/16.A, para. 61, (Jan. 15, 2018).

²¹See *MM (DRC – Plausibility) Democratic Republic of Congo v. Secretary of State for the Home Department*, [2005] UKIAT 00019, para. 26 (Eng.).

²²See VG Berlin, Case No. 8 K 794.16 A, para. 31; see also HUBERT HEINOLD, RECHT FÜR FLÜCHTLINGE 109 f. (7th ed. 2015).

²³See ECJ, Cases C-148/13 and C-150/13, A, B and C, A et al. v. Staatssecretaris van Veiligheid en Justitie, ECLI:EU:C:2014:2406 (Dec. 2, 2014), paras. 49–53; Opinion of Advocate General Sharpston, paras. 36–49, Cases C-148/13, C-149/13 and C-150/13, A, B and C, ECLI:EU:C:2014:2111 (July 17, 2014).

²⁴See, e.g., Uwe Berlit, Harald Doerig & Hugo Storey, *Religious Conversion and Homosexuality*, 27 INT’L. J. REFUGEE L. 649 (2015).

²⁵See HUNGARIAN HELSINKI COMMITTEE, CREDIBILITY ASSESSMENT IN ASYLUM PROCEDURES – A MULTIDISCIPLINARY TRAINING MANUAL, Vol. 2, 74 (2015).

²⁶See EASO, *Evidence and Credibility Assessment in the Context of the Common European Asylum System* 180 (2018), https://euaa.europa.eu/sites/default/files/EASO%20Evidence%20and%20Credibility%20Assesment_JA_EN_0.pdf; UNHCR, Guideline on International Protection No. 9: Sexual Orientation/Gender Identity para. 62 (2012), <https://www.unhcr.org/509136ca9.pdf>.

²⁷See HUNGARIAN HELSINKI COMMITTEE, *supra* note 25, at 77. See also Jenni Millbank, *The Ring of Truth: A Case Study of Credibility Assessment in Particular Social Group Refugee Determinations*, 21 INT’L J. REFUGEE L. 1, 17 (2009) (detailing a similar model in Canada).

further acts of persecution or by claiming more serious versions of the original account later in the proceedings.²⁸ Because it is *prima facie* implausible why these new facts were not advanced earlier, this is taken to be an indicator against the account's credibility.

All of these criteria are accepted internationally, for example by the UNHCR,²⁹ EASO,³⁰ the International Association of Refugee Law Judges (IARLJ),³¹ and the European Court of Human Rights (ECtHR).³² But the literature often splits these criteria into many more categories. *Bender et al.*, the German standard reference work, distinguishes between criteria relating to the content and structure of the account, and further divides these into many more categories.³³ *Aldert Vrij* mentions no less than nineteen criteria.³⁴ But all of the criteria that are mentioned in this literature, and relevant for asylum law, can be covered by the simplified categorization advocated for here. Further differences could be made within these categories if need be. The qualitative analysis already undertaken suggests that this simplified version better reflects how credibility criteria are used in asylum law practice.³⁵ The quantitative analysis undertaken here also served to test the adequacy of this systematization by considering whether additional (sub-)categories were necessary to adequately take into account the practice of credibility assessment.

2. Conduct-Based Criteria

In addition to the generally accepted content-based credibility criteria, conduct-based criteria are sometimes used: The conduct of a person during the asylum interview, before the Federal Agency or at court,³⁶ and the applicant's personal or general credibility.

First, the *conduct* of the applicant during an *asylum interview* concerns communicative behavior beyond the meaning of the words that are spoken. This demeanor can be non-verbal, as avoiding eye contact, but it can also be the way something is said, e.g. long pauses between sentences or the volume and speed of speech.³⁷ This criterion has widely and for a long time been criticized as unreliable, a critique which will be developed in more detail below. Nonetheless, it is an accepted

²⁸See Bundesverwaltungsgericht [BVerwG] [Federal Administrative Court], Apr. 16, 1985, ENTSCHIEDUNGEN DES BUNDESVERWALTUNGSGERICHTS [BVerwGE] 71, 180–183, para. 18 (“[B]ei erheblichen Widersprüchen oder Steigerungen im Sachvortrag [kann] dem Kläger nur bei einer überzeugenden Auflösung der Unstimmigkeiten geglaubt werden.”). See also BERLIT ET AL., *supra* note 24, at 652; RALPH GÖBEL-ZIMMERMANN, ALEXANDER EICHHORN & STEPHAN BEICHEL-BENEDETTI, ASYL- UND FLÜCHTLINGSRECHT, paras. 482, 690 (2017).

²⁹See UNHCR, Beyond Proof: Credibility Assessment in EU Asylum Systems at 260 (May 2013), <https://www.unhcr.org/51a8a08a9.pdf>; EASO, *supra* note 26, at 84.

³⁰See EASO, *supra* note 26, at 85.

³¹See IARLJ, *supra* note 15, at 33 f.; BERLIT ET AL., *supra* note 24, at 652; HUNGARIAN HELSINKI COMMITTEE, CREDIBILITY ASSESSMENT IN ASYLUM PROCEDURES – A MULTIDISCIPLINARY TRAINING MANUAL, Vol. 1, 32 (2013).

³²See Muradi & Alieva v. Sweden, App. No. 11243/13, para. 44 f. (Jun. 25, 2013); Tekdemir v. the Netherlands, App. No. 49823/99 and 46860/99 (Oct. 1, 2002); Hakizimana v. Sweden, App. No. 37913/05 (Mar. 27, 2008); M.K.N. v. Sweden, App. No. 72413/10, para. 43 (Jun. 27, 2013); R.C. v. Sweden, App. No. 41827/07, para. 44 (Mar. 9, 2010).

³³See BENDER ET AL., *supra* note 7, at paras. 424 ff. (including such categories as level of detail, conversations, complications, interactions, originality, emotions, pace, homogeneity and much more).

³⁴See Vrij, *supra* note 5, at 209–212. The criteria are: (1) Logical Structure, (2) Unstructured Production, (3) Quantity of Detail, (4) Contextual Embedding, (5) Descriptions of Interactions, (6) Reproduction of Conversation, (7) Unexpected Complications during the Incident, (8) Unusual Details, (9) Superfluous Details, (10) Accurately Reported Details Misunderstood, (11) Related External Associations, (12) Accounts of Subjective Mental State, (13) Attribution of Perpetrator's Mental State, (14) Spontaneous Corrections, (15) Admitting Lack of Memory, (16) Raising Doubts About One's Own Testimony, (17) Self-Deprecation, (18) Pardoning the Perpetrator, (19) Details Characteristics of the Offence.

³⁵BJÖRNSTJERN BAADE, WAHRHEIT UND RECHT – STÖRUNG UND SCHUTZ REGULATORISCHER, ASYLRECHTLICHER UND MEDIALER WAHRHEITSFINDUNG 233 et seq. (forthcoming); see also HUNGARIAN HELSINKI COMMITTEE, *supra* note 31, at 32.

³⁶Legally, there are technical differences between the asylum interview before the Federal Agency, conducted according to Section 24 Asylum Law, and the hearing before a court, in which the applicant is in practice “informally heard” (informatorische Anhörung). For the purposes of credibility assessment, no relevant discrepancies result from these formal differences.

³⁷See BENDER ET AL., *supra* note 7, paras. 479 et seq.

credibility criterion, not only in asylum cases.³⁸ In the USA, the law expressly recognizes that a person's "demeanor, candor, or responsiveness" can be taken into account in asylum cases.³⁹ The ECtHR accepts it as well. Much as US federal courts,⁴⁰ the ECtHR sees it as a reason to widen the margin of appreciation that national authorities enjoy because only they perceived this conduct.⁴¹

Second, German theory distinguishes between *the account's* credibility (*Glaubhaftigkeit*) and *a person's* credibility (*Glaubwürdigkeit*).⁴² Thus, not only the account's credibility but the credibility of the person telling it is a criterion used in German asylum law, and in other branches of the law. Widely criticized, this criterion is nonetheless generally accepted as well, in particular by the German Federal Constitutional Court.⁴³ Other national courts, e.g. in Canada⁴⁴ and the UK,⁴⁵ have accepted it, just as the ECtHR which refers to it as "general" or "personal credibility."⁴⁶ Article 4 (5) Qualification Directive (QD)⁴⁷ likewise refers to "general credibility."

II. Confounding Factors

Anyone conducting a credibility assessment must be aware of confounding factors. Some of these confounding factors are general and applicable to credibility assessment in all areas of law, others are rather specific to asylum cases. For example, memory may fail anyone, especially concerning

³⁸See EASO, *supra* note 26, at 90; REINHARD MARX, ASYL: KOMMENTAR ZUM ASYLGESETZ, § 24, para. 22 (10th ed. 2019). On the great importance in U.S. criminal justice, contrary to academic objections, see Max Minzner, *Detecting Lies Using Demeanor, Bias, and Context*, 29 CARDOZO L. REV. 2557 (2008); Aldert Vrij, *Nonverbal Detection of Deception, in FINDING TRUTH IN THE COURTROOM: DEALING WITH DECEPTION, LIES, AND MEMORIES* 174 (Henry Otgaar & Mark L. Howe eds., 2018); Millbank, *supra* note 27, para. 6 f. See also Max Hermanutz & Sven Litzcke, *Polizeiliche Vernehmung, in VERNEHMUNG IN THEORIE UND PRAXIS: WAHRHEIT-IRRTUM-LÜGE* 71 (Max Hermanutz & Sven Litzcke eds., 3rd ed. 2012).

³⁹Immigration and Nationality Act § 208(b)(1)(B)(iii), 8 U.S.C. § 1158(b)(1)(B)(iii).

⁴⁰See *Singh-Kaur v. Immigr. & Naturalization Serv.*, 183 F.3d 1147, 1151 (1999) ("special deference").

⁴¹R.C. v. Sweden, App. No. 41827/07, para. 52, <https://hudoc.echr.coe.int/fre?i=002-1023> ("The Court . . . accepts that, as a general principle, the national authorities are best placed to assess not just the facts but, more particularly, the credibility of witnesses since it is they who have had an opportunity to see, hear and assess the demeanour of the individual concerned."); cf. Opinion of Advocate General Sharpston, *supra* note 23, para. 73 ("It is therefore important that the official making the determination has an opportunity to see the applicant giving his account or at the very least has a full report as to his demeanour during the course of the examination (my preference is for the former).").

⁴²See, e.g., Bundesverfassungsgericht [BVerfG] [Federal Constitutional Court], May 15, 1996, 94 ENTSCHEIDUNGEN DES BUNDESVERFASSUNGSGERICHTS [BVerfGE] 166, 205.

⁴³See Bundesverfassungsgericht [BVerfG] [Federal Constitutional Court], May 14, 1996, 94 ENTSCHEIDUNGEN DES BUNDESVERFASSUNGSGERICHTS [BVerfGE] 115, 147; Bundesverfassungsgericht [BVerfG] [Federal Constitutional Court], May 15, 1996, 94 ENTSCHEIDUNGEN DES BUNDESVERFASSUNGSGERICHTS [BVerfGE] 166, 201.

⁴⁴*Sheikh v. Canada* (Minister of Employment and Immigration), 1990 CanLII 8017 (FCA), para. 8:

I would add that in my view, even without disbelieving every word an applicant has uttered, a first-level panel may reasonably find him so lacking in credibility that it concludes there is no credible evidence relevant to his claim on which a second-level panel could uphold that claim. In other words, a general finding of a lack of credibility on the part of the applicant may conceivably extend to all relevant evidence emanating from his testimony. Of course, since an applicant has to establish that all the elements of the definition of Convention refugee are verified in his case, a first-level panel's conclusion that there is no credible basis for any element of his claim is sufficient.

⁴⁵See UK Home Office, *Asylum Policy Instruction: Assessing Credibility and Refugee Status* 13 (Jan. 6, 2015), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/397778/ASSESSING_CREDIBILITY_AND_REFUGEE_STATUS_V9_0.pdf; Anthony Good, 'The Benefit of Doubt' in *British Asylum Claims and International Cricket*, in *OF DOUBT AND PROOF: RITUAL AND LEGAL PRACTICES OF JUDGMENT* 124 (Daniela Berti, Anthony Good & Gilles Tarabout eds., 2015).

⁴⁶*Collins & Akaziebie v. Sweden*, App. No. 23944/05 (Mar. 8, 2007), <https://hudoc.echr.coe.int/eng?i=001-79864>.

⁴⁷Council Directive 2011/95/EU, art. 9–26, 2011 O.J. (L 337) (on standards for the qualification of third-country nationals or stateless persons as beneficiaries of international protection, for a uniform status for refugees or for persons eligible for subsidiary protection, and for the content of the protection granted (recast)).

the details of an event after some time has passed.⁴⁸ Even more, our memory is never a photo-realistic copy of reality but focuses on the perceptions that seemed important to us at the time.⁴⁹ Therefore, there is no general rule *which* details of a situation must be remembered in a credible account; age, gender, education, and other aspects of one's socialization may all have an impact.⁵⁰ While crucial events in our lives will generally engrave themselves deeper in our memory,⁵¹ someone who suffers from post-traumatic stress disorder might present an account that is internally inconsistent or exhibits other negative credibility criteria.⁵²

Mistakes by interpreters, whose assistance is almost always necessary in asylum cases, can likewise create the appearance of negative credibility criteria, e.g. inconsistencies or a lack of detail.⁵³ Such mistakes can be hard to detect because usually⁵⁴ only one of the persons present at the interview speaks both languages: The interpreter.⁵⁵

Cultural distance between the applicant and the decision-maker can be an issue for credibility determination, too.⁵⁶ What kind of "story-telling" is perceived as credible can differ between cultures, e.g. how emotionally or directly an event is recounted.⁵⁷ It can be more common in one culture to give additional details about one's own experience and emotions, and in another culture the focus may lie on details about social interactions; it may be normal to give more or less context. All of this may affect the level of detail of the account—and also what is remembered because it seemed important at the time.⁵⁸ Specific dates might be less important in some cultures and therefore less well remembered.⁵⁹ It is a problem often described in the literature on credibility determination in asylum cases that the applicant who wants to tell something that is important from their point of view is interrupted because the interviewer considers it to be irrelevant.⁶⁰ International criminal courts have noticed this problem as well: The International Criminal Tribunal for Rwanda explicitly acknowledged that questions are not always answered directly in Rwanda, specifically when they concern sensitive issues.⁶¹ The demeanor during an interview can likewise be culturally contingent, in particular gestures and facial expressions.⁶² Looking down, for example, is considered to be a sign of respect towards authorities in Hispanic cultures—but is often taken as a "cue" for lying in other Western cultures.⁶³

⁴⁸Juliet Cohen, *Questions of Credibility: Omissions, Discrepancies and Errors of Recall in the Testimony of Asylum Seekers*, 13 INT'L J. REFUGEE L. 293, 295 (2001).

⁴⁹See HUNGARIAN HELSINKI COMMITTEE, *supra* note 31, at 65–67.

⁵⁰See HUNGARIAN HELSINKI COMMITTEE, *supra* note 25, at 42–43, 99–100.

⁵¹See Cohen, *supra* note 48, at 295–296, 298.

⁵²See Cohen, *supra* note 48, at 303; BAFF, *supra* note 12, at 60–61.

⁵³Cf. Sarah Craig & Karin Zwaan, *Legal Overview*, in ASYLUM DETERMINATION IN EUROPE 36 f. (Nick Gill & Anthony Good eds., 2019); ANTHONY GOOD, ANTHROPOLOGY AND EXPERTISE IN ASYLUM COURTS 185–186 (2006).

⁵⁴According to § 17(2) GK-AsylG, applicants may be accompanied by an interpreter of their own choosing (but they must bear any costs).

⁵⁵SUSAN BERK-SELIGSON, *THE BILINGUAL COURTROOM: COURT INTERPRETERS IN THE JUDICIAL PROCESS* 197 (2nd ed. 2017).

⁵⁶UNHCR, Note on Determination of Refugee Status under International Instruments, EC/SCP/5 (Aug. 24, 1977), para. 13.

⁵⁷See HUNGARIAN HELSINKI COMMITTEE, *supra* note 31, at 115–117.

⁵⁸See EASO, *supra* note 26, at 176; HUNGARIAN HELSINKI COMMITTEE, *supra* note 31, at 76–77, 119–220; BRIGITTE BRAND & JÖRG WEIDENHAMMER, BEDINGUNGEN FÜR DIE ANHÖRUNG VON FLÜCHTLINGEN IM RAHMEN DES ASYLVERFAHRENS, DIE UNTER ANDEREM FOLTER UND ANDERE TRAUMATISCHE ERFAHRUNGEN ERLITTEN HABEN 26 (1991).

⁵⁹Walter Kälin, *Troubled Communication: Cross-Cultural Misunderstandings in the Asylum-Hearing*, INT'L MIGRATION REV. 230, 236–237 (1986).

⁶⁰*Id.* at 232.

⁶¹See Prosecutor v. Akayesu, Case No. ICTR-96-4-T, Judgment, § 156 (Sep. 2, 1998); Robert Cryer, *A Message from Elsewhere: Witnesses Before International Criminal Tribunals*, in INNOVATIONS IN EVIDENCE AND PROOF 390–391 (Paul Roberts & Mike Redmayne eds., 2007).

⁶²See Thomas Bingham, *The Judge as Juror: The Judicial Determination of Factual Issues*, in CURRENT LEGAL PROBLEMS 1, 10–11 (1985).

⁶³See VRIJ, *supra* note 38, at 165; BERK-SELIGSON, *supra* note 55, at 254; Cryer, *supra* note 61, at 393.

Confounding factors are thus a substantial challenge in credibility determination. But the mere possibility that distortive factors could generally have an influence on credibility assessment does not per se invalidate the credibility criteria, as is sometimes argued.⁶⁴ Just because confounding factors *could*, abstractly speaking, be present in any case, that does not mean that they are present in all cases. Nor does it mean that it is impossible to know to a reasonable degree of certainty in which cases they are present. It also does not mean that it is impossible to mitigate them or know the limits of their impact. For example, if it is reasonable to assume that dates cannot be remembered well, it can be asked instead if something happened before or after a key event, e.g. a wedding.⁶⁵ Even traumatic events do not always lead to memory loss,⁶⁶ and the periphery of a traumatic event, what happens before or after it, can generally be remembered.⁶⁷

III. Balancing and the Purpose of the Asylum Interview

Credibility assessment is a very complex balancing operation. The criteria can initially be ambivalent and contradictory. To arrive at a decision, the criteria therefore need to be clarified and balanced against each other, together with possible confounding factors—which is often described as the need to consider all criteria and confounding factors “in the round.”⁶⁸ The process of carefully balancing credibility criteria, while taking into account distortive factors, is what differentiates this type of content-based credibility assessment from common beliefs about “lie detection” that even professional decision-makers are prone to succumb to. The existence of any one negative credibility criterion must not be overvalued. For example, a slight contradiction cannot in itself make the entire account incredible.⁶⁹

With regard to the balancing thus required, the asylum interview before the Federal Agency or at court serves three central purposes. *First*, it generates the account and thus the object of credibility assessment. *Second*, it is supposed to generate further and sufficiently clear credibility criteria to determine whether the account is credible or not. Abstractly, the credibility criteria can seem ambivalent: Consistency, relevant knowledge, and a high level of detail are valued as positive criteria, but too much consistency, too many details and a certain kind of knowledge can make the impression of an invented, rehearsed story.⁷⁰ In the interview, the criteria should become more unequivocal, or should be supplemented by additional indicators that enable a well-founded decision. *Third*, during the interview, decision-makers need to clarify whether there is reason to believe that negative credibility criteria are based on *confounding factors*—and therefore do not weigh against the account’s credibility. For example, an inconsistency in the account might be explained by a simple translation error.

⁶⁴See, e.g., Noll, *supra* note 4, at 616.

⁶⁵See HUNGARIAN HELSINKI COMMITTEE, *supra* note 31, at 74–75.

⁶⁶*Id.* at 97.

⁶⁷Lars Dittrich & Barbara Wild, *Über den Umgang mit Traumatisierungen im Asylprozess*, in NEUE ZEITSCHRIFT FÜR VERWALTUNGSRECHT [NVwZ] 118, 123 (2020).

⁶⁸See IARLJ, *supra* note 15, at 33; EASO, *supra* note 26, at 84; UNHCR, Note on Burden and Standard of Proof (Dec. 16, 1998), para. 11; cf. Sean Rehaag & Hilary Evans Cameron, *Experimenting with Credibility in Refugee Adjudication: Gaydar*, CANADIAN J. OF HUM. RTS. 1, 17–21 (2020); UK Home Office, *supra* note 45, at 13–14.

⁶⁹See Bundesverwaltungsgericht [BVerwGE] [Federal Administrative Court], Apr. 16, 1985, 71 ENTSCHIEDUNGEN DES BUNDESVERWALTUNGSGERICHTS [BVerwGE] 180–183, para. 18; CHRISTOPH GUSY, DAS ASYLRECHT IN DER RECHTSPRECHUNG DES BUNDESVERFASSUNGSGERICHTS 61 (1983). *But see* Bundesverfassungsgericht [BVerfG] [Federal Constitutional Court], Jun. 30, 1964, 18 ENTSCHIEDUNGEN DES BUNDESVERFASSUNGSGERICHTS [BVerfGE] 112, 115 (representing an essential contradiction by all means).

⁷⁰See EASO, *supra* note 26, at 86; Cohen, *supra* note 48, at 308; GOOD, *supra* note 53, at 171; BERLIT ET AL., *supra* note 24, at 658. See, e.g., Verwaltungsgericht Ansbach [VG] [Administrative Trial Court of Ansbach], Case No. AN 3 K 12.30067, at 11, (Apr. 24, 2012) (“The claimant has merely uttered empty phrases that anyone can read up in generally accessible sources or learn in courses.”) (translation by the authors).

C. Empirical Research on Credibility Criteria So Far

There is no official data on the application of credibility criteria in Germany. The Federal Agency publishes annual statistics, but they just contain general information such as the number of cases, the applicants' countries of origin, and their success rate. Not even the reasons for which asylum was claimed are documented statistically.⁷¹ The statistics that are published on a European level by EASO do not relate to credibility criteria either.⁷²

In Germany, case law is rarely studied quantitatively.⁷³ But it has been found that recognition rates can vary significantly between individual decision-makers.⁷⁴ It has also been shown that acceptance rates vary significantly between the German states (*Länder*)—which some call the “dark side” of federalism, because applicants cannot choose in which state their case will be heard.⁷⁵ The general atmosphere with regard to refugees was taken to have an influence on whether an application is granted or rejected.⁷⁶ One study even found that in German states long governed by the social democrats of the SPD rejection rates are lower in administrative courts, indicating—according to the authors—that “partisanship is a factor to be reckoned with.”⁷⁷ A further study found that males, Muslims, and persons whose status determination was conducted in more conservative areas of Germany were less likely to receive a protective status. The study considered this to reveal “taste-based discrimination,” i.e. a decision based on extra-legal reasons, such as stereotypes and prejudice. Such decisions were seen to be more prevalent among decision-makers under a high workload or who disposed of little information.⁷⁸ But specifically with regard to their credibility assessment, German asylum decisions have not been quantitatively analyzed thus far, neither the decisions by the Federal Agency nor those by the administrative courts.

While a quantitative analysis of decision-making has been more prominently conducted for the case law and voting behavior of judges at constitutional courts,⁷⁹ some academic literature outside of Germany has begun to explore the nature and conduct of credibility assessment in asylum cases. In particular, members of York University's Centre for Refugee Studies in Toronto, Canada, pioneered research on this issue.

Sean Rehaag and *Hilary Evans Cameron* investigated the impact of prejudice on credibility assessment in an experiment. They showed a fictitious asylum case file to 284 Canadian first-year law students. The students' task was to determine the credibility of the fictional applicant who claimed to be homosexual. The case file included positive as well as negative credibility criteria,

⁷¹See Bundesministerium für Familie, Senioren, Frauen und Jugend [Federal Ministry for Family Affairs, Senior Citizens, Women and Youth], Group of Experts on Action against Violence against Women and Domestic Violence (GREVIO): Erster Staatenbericht der Bundesrepublik Deutschland 69, <https://www.bmfsfj.de/resource/blob/160138/6ba3694cae22e5c9af6645f7d743d585/grevio-staatenbericht-2020-data.pdf> (2020) (“Das BAMF erfasst die Asylgründe nicht statistisch.”).

⁷²See European Union Agency for Asylum (EUAA), Data Analysis and Research <https://www.easo.europa.eu/analysis-and-statistics> (last visited Apr. 10, 2022).

⁷³See Corinna Coupette & Andreas M. Fleckner, *Quantitative Rechtswissenschaft: Sammlung, Analyse und Kommunikation juristischer Daten*, in *JURISTENZEITUNG [JZ]* 379, 389 (2018). But see Sebastian J. Goerg & Niels Petersen, *Empirische Methoden*, in *ÖKONOMISCHE METHODEN IM RECHT* 195–235 (Niels Petersen & Emanuel V. Towfigh eds., 2nd ed. 2017) (exemplifying those who pioneer empirical methods).

⁷⁴In Germany, but also in Canada and the USA. See Noll, *supra* note 4, at 618–620.

⁷⁵Gerald Schneider, Nadine Segadlo & Miriam Leue, *Forty-Eight Shades of Germany: Positive and Negative Discrimination in Federal Asylum Decision Making*, 29 GER. POL. 564, 568569 (2020).

⁷⁶*Cf. id.* at 573, finding that higher concern for security in a German state, indicated by higher spending in this area, correlates with higher rejections, just like higher unemployment does.

⁷⁷*Id.* at 575.

⁷⁸Lidwina Gundacker, Yuliya Kosyakova & Gerald Schneider, *Global Norms, Regional Practices: Taste-Based & Statistical Discrimination in German Asylum Decision-Making* (University of Konstanz, Cluster of Excellence “The Politics of Inequality,” Working Paper No. 5, 2021), <https://kops.uni-konstanz.de/server/api/core/bitstreams/e86db1fb-cba7-49db-9707-85eb0e53e6c6/content>.

⁷⁹See generally JEFFREY A. SEGAL & HAROLD J. SPAETH, *THE SUPREME COURT AND THE ATTITUDINAL MODEL REVISITED* (2007).

so as to make it possible to decide either way. One group of students was explicitly advised not to take into account stereotypes concerning the applicant's appearance, the other was not. Both groups were again divided into three sub-groups, which were shown either no photo of the applicant, a photo that showed someone who conformed to Western stereotypes of a homosexual person or a photo that did not do so.

Despite certain limitations,⁸⁰ the experiment yielded interesting insights. Only one person explicitly referred to the applicant's looks, which they took to support the claim. Thus, almost all students knew that they should not base their decision on this factor—regardless of whether they belonged to the group that had received advice on this issue or not. The group that saw a photo that conformed to homosexual stereotypes was nonetheless more likely to consider the applicant credible (88 % of this group did) than those who saw a photo meant to be stereotypically “heterosexual” (only 75.5 % of this group did). 85.3 % of the control group with no photo considered the applicant credible.⁸¹

Sean Rehaag also researched the infamous case of the Canadian decision-maker *McBean*. In three years, *McBean* rejected all 174 asylum claims that he decided. Of all these applicants, 116 were denied refugee status for a lack of credibility. Pointing to a lack of details and inconsistencies, often with regard to dates or the exact number of attackers during an act of persecution, the decision-maker without fail came to the result that “I simply do not believe . . .”⁸² The likelihood that all these asylum claims were indeed unfounded is close to zero.⁸³

Finally, *Jenni Millbank* of the University of Technology in Sydney, Australia, analyzed 1,000 credibility determinations in asylum cases decided in Australia, the United Kingdom, Canada, and New Zealand concerning the applicants' sexual orientation.⁸⁴ In particular, the 149 Canadian cases showed that legal counsel can make a difference: Those applicants who were not represented by counsel were successful in only 2.3 % of cases; those represented by counsel were successful in 29 % of cases. 56 % of applicants did not have a lawyer.⁸⁵

D. An Empirical Study of Credibility Assessment

Inspired by the research done in this area so far, this article seeks to understand how credibility assessment is conducted in the practice of German courts using a quantitative approach. It will seek to ascertain whether and how the framework of credibility assessment that the qualitative research revealed is operationalized in practice. The aim of this study is to analyze the use of credibility criteria and the way that confounding factors are taken into account.

First, the study's preparation and conduct, including its limitations will be described (I.), then the results will be presented descriptively (II.) before they are analyzed (III.).

⁸⁰The results of this experiment underlie certain limitations because it was conducted in connection with a second experiment on credibility determination that might have influenced the first one. The students might have fallen for the “gambler's fallacy” because those who found the one applicant credible were more likely to find the other one not credible. See Rehaag & Cameron, *supra* note 68, at 7–12.

⁸¹See Rehaag & Cameron, *supra* note 68, at 13.

⁸²Sean Rehaag, *I Simply do not Believe . . . : A Case Study of Credibility Determinations in Canadian Refugee Adjudication*, 38 WINDSOR REV. OF L. & SOC. ISSUES 38, 49–51 (2017).

⁸³See *id.* (providing the calculation of this probability).

⁸⁴See Millbank, *supra* note 27, at 2–3.

⁸⁵See Millbank, *supra* note 27, at 23.

1. Preparation and Conduct of the Study⁸⁶

1. Purpose and Method of the Study

This quantitative analysis is meant to paint a more complete picture of the credibility assessment actually conducted in German courts.⁸⁷ Which criteria are relied on the most in practice? How does this application of reality criteria relate to evidentiary issues that are often discussed in the literature? Qualitative analyses are usually restricted to an impressionistic evaluation of quantities. Based on the analysis of a limited number of cases, it is common to postulate that certain issues come up “often,” others maybe “rarely.” While this impression need not be wrong, an empirical approach may serve to confirm such intuitions.

Additionally, this analysis may be able to inquire—to a certain extent—whether a relationship exists between the application of certain credibility criteria and confounding factors on the one hand and other properties of the cases, such as the applicant’s gender or country of origin. Due to the limited number of cases and the broader study design, such determinations are likely to be of limited reliability. But as a first study that is partly exploratory, partly descriptive, it may nonetheless show where additional research may be necessary.

In order to pursue these goals, a sample of first-instance administrative court decisions was analyzed with regard to credibility criteria, confounding factors, certain evidentiary issues and further potentially relevant properties.

First, we recorded some general properties: The court and date of the decision, the case number, the type of decision,⁸⁸ the applicant’s country of origin, gender, and the kind of persecution that was claimed.

Second, it was recorded whether the court found the applicant’s account credible or incredible, and when a court based its negative decision on the applicant’s burden of proof. It was also recorded when the court found that it was not necessary to decide on credibility because a decision could be reached without making that determination.

Third, the use of credibility criteria was recorded and categorized according to the qualitative assessment explained above: *Consistency*, *plausibility*, *personal credibility*, and *conduct during the interview*, including the type or aspect of the applicant’s conduct that was held to be relevant, e.g. emotion or aggression.

These general criteria were divided into further sub-categories: Plausibility was divided into *general plausibility*, the *level of detail*, *knowledge*, and the *timeliness of the claim*. Consistency was divided into *internal*, and *external consistency*. External consistency was further subdivided into the various means of proof that could, according to the qualitative literature, play a role in asylum cases: *Country-of-origin information (COI)*, *witnesses*, *medical expert opinions* (commissioned by the applicant or the court), *country expert opinions*, *data from mobile phones*, *reports in the media*, *documents* (private or official), and *judicial inspection (Augenschein)* e.g. of photos or the applicant’s body.

Fourth, we recorded whether and how courts took into account as confounding factors: Age, gender, psychological strain, eloquence, mistakes in interview transcripts or mistakes of interpreters, cultural distance, simple error, or other confounding factors.

Fifth, it was recorded whether and to what extent the courts relied on procedural mechanisms that are particular to asylum cases: Article 4 (5) QD, which lightens the burden of proof for the applicant under certain conditions, and Section 77 (2) Asylum Act (*Asylgesetz—AsylG*),⁸⁹ which

⁸⁶Cf. Ashutosh Avinash Bhagwat & Matthew Struhar, *Justice Kennedy’s Free Speech Jurisprudence: A Quantitative and Qualitative Analysis*, 44 MCGEORGE L. REV. 167 (2014) (generally detailing the approach).

⁸⁷Cf. Urška Šadl & Henrik Palmer Olsen, *Can Quantitative Methods Complement Doctrinal Legal Studies? Using Citation Network and Corpus Linguistic Analysis to Understand International Courts*, 30 LEIDEN J. INT’L L. 327, 330 (2017) (for the usefulness of a quantitative decision evaluation).

⁸⁸Judgment, decision, or summary decision (“Gerichtsbescheid”) according to Section 84 of the Code of Administrative Court Procedure.

⁸⁹2008 (Federal Law Gazette I at 1798), most recently amended by Article 1 of the Act of Dec. 21, 2022 (Federal Law Gazette I at 2817), English translation (as of 2016) available at: https://www.gesetze-im-internet.de/englisch_asylvfg/index.html.

allows a court to dispense with a further presentation of the facts and of the reasons for its decision, provided that it follows the statements and justification of the Federal Agency.

2. Case Selection and Overview of Cases

In Germany, credibility assessment is first conducted by the executive decision-makers of the Federal Agency who decide on the asylum claim as an administrative agency. The applicant can have a negative decision reviewed by the first-instance administrative courts and ask them to issue a positive decision, Section 113 (V) of the Code of Administrative Court Procedure (*Verwaltungsgerichtsordnung—VwGO*).⁹⁰ These courts will conduct the credibility assessment anew. This decision by the first-instance courts will in most cases be the final one before the administrative courts, due to various restrictions on the right to appeal in German asylum law.⁹¹

Many of these first-instance court decisions are available in official databases. Decisions of the Federal Agency are mostly not publicly available. This study will therefore focus on first-instance administrative court decisions. According to the Federal Agency, in 2017, German administrative courts took 158,726 decisions in asylum cases. Of these cases, 92.6 % (146,168) were decisions by first-instance administrative courts.⁹²

Due to limited resources, not all cases available in the database could be analyzed. A timeframe of three months was therefore chosen to select a manageable number of cases. The timeframe could not be designated in an entirely random manner.⁹³ The year 2017 was chosen because this was the first year following the “migration crisis” of 2015, in which the number of asylum cases decided by German administrative courts rose starkly and the rate at which the Federal Agency’s decisions were invalidated was the highest: More than 140,000 cases were decided in 2017 (2016: about 70,000), and 22 % of Federal Agency decisions were invalidated in 2017 (2016: 13.1 %).⁹⁴ This promised many relevant decisions that concerned credibility criteria and confounding factors. To keep the scope of this first inquiry manageable, only the months from January to March were chosen. Within the timeframe thus chosen, no further selections were made.

All first-instance administrative court decisions which had been decided in the first quarter of 2017, contained the terms “Flüchtlingseigenschaft” (refugee status) and “glaubhaft” (credible) were selected. A more accurate selection was not possible with the case properties available in the database.⁹⁵ The ambiguity of the German term “glaubhaft” was problematic for case selection. It refers not only to the credibility assessment under investigation here but also to a standard of proof.⁹⁶

The cases were downloaded from the case law database Juris. This database contains a small part of all court decisions that are actually taken.⁹⁷ According to a 2021 study by *Hanjo Hamann*, less than 1 % of all criminal and civil decisions by German courts are published annually.⁹⁸

⁹⁰1991 (Federal Law Gazette I at 686), most recently amended by Article 3 of the Act of Jul. 20, 2022 (Federal Law Gazette I at 1325). English translation (as of 2019) available at: https://www.gesetze-im-internet.de/englisch_vwgo/englisch_vwgo.html.

⁹¹See in particular § 78 Asylum Act. A complaint to the Federal Constitutional Court or the European Court of Human Rights is possible, but generally these courts will not conduct a new credibility assessment. Normally, they will only declare that arbitrary credibility assessments violate the constitution or the Convention.

⁹²FEDERAL AGENCY, DAS BUNDESAMT IN ZAHLEN 2017: ASYL, MIGRATION UND INTEGRATION 60 (2018).

⁹³Cf. PETER CANE & HERBERT M. KRITZER, THE OXFORD HANDBOOK OF EMPIRICAL LEGAL RESEARCH 36, 53 (2021) (explaining that a sample that is formally random will not always be available, but meaningful results can be gained despite this).

⁹⁴Federal Agency, Gerichtsstatistik 2019, Mar. 30, 2020, <https://www.bamf.de/SharedDocs/Meldungen/DE/2020/20200330-am-gerichtsstatistik-2019.html?nn=282388>.

⁹⁵The same limitation applies to the largest commercial competitor, the database BECK ONLINE, www.beck-online.de (last visited Apr. 14, 2022).

⁹⁶See ZIVILPROZESSORDNUNG [ZPO] [CODE OF CIVIL PROCEDURE], § 294 (Ger.). This standard of proof is the functional equivalent of the “preponderance of evidence” in common law jurisdictions.

⁹⁷See Coupette & Fleckner, *supra* note 73, at 380–381. See also Millbank, *supra* note 27, 3, para. 13.

⁹⁸See Hanjo Hamann, *Der blinde Fleck der deutschen Rechtswissenschaft – Zur digitalen Verfügbarkeit instanzgerichtlicher Rechtsprechung*, 76 JURISTENZEITUNG 656, 658 (2021).

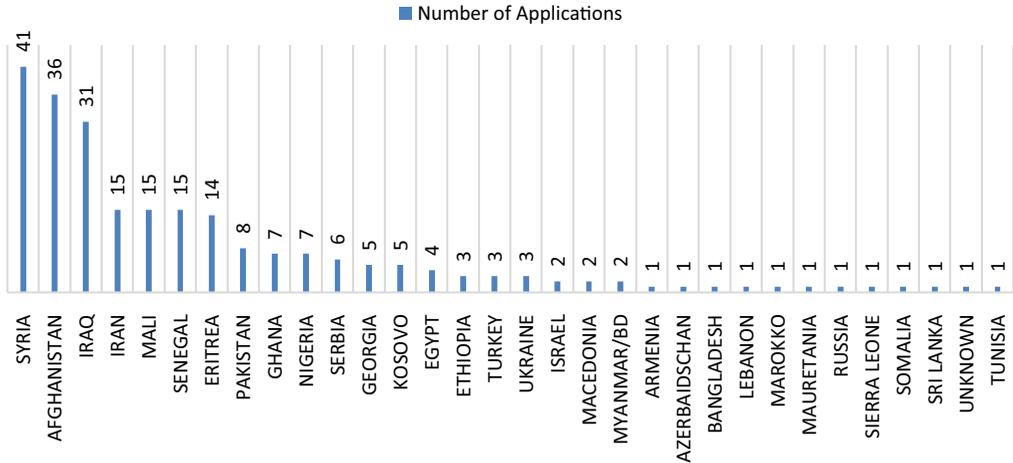


Figure 1. Countries of Origin in the Sample.

But Juris is the largest case law database in Germany. It is operated by a private-law company that is majority-owned by the German state, and cooperates with the courts. There is no general obligation or practice of courts to submit all of their decisions for publication to Juris. Cases that the courts themselves do not deem “worthy of publication,” i.e. of sufficient legal interest, will usually not be included in a database.⁹⁹ For these reasons, the sample could not be representative of all decisions made.

Empirically analyzing the case law available on Juris can provide important insights nonetheless. *First*, Juris is used by basically all courts and practitioners. The cases available on Juris thus influence the way that practitioners see and discuss the law. *Second*, the study uses the data that is actually available to see and assess much more of the case law than is commonly done in legal research, and thus achieves a broader and more objective impression of that case law. While the sample could not be entirely representative, it achieves an overview that is more representative of the case law than that achieved by the common practice of using merely a handful of cases, selected from this same available case law.

This case selection yielded a pool of 291 decisions.¹⁰⁰ Of these 291 first-instance administrative court decisions, 55 turned out not to be relevant. In these 55 cases, the courts did not decide on the applicant’s refugee status. 236 decisions accordingly remained to be analyzed with regard to their handling of the credibility assessment.

The sample includes a wide range of countries of origin. It is not representative of the total number of asylum cases decided by first-instance courts in 2017, but the sample includes cases from all top ten countries of origin in that year, except for Albania (Figs. 1 and 2).

Most applicants in the sample were male (64%). Couples or single parents with children (17%) and female applicants (13%) were likewise represented. Only one applicant identified as non-

⁹⁹See Hamann, *supra* note 97, at 659.

¹⁰⁰The database query resulted in 294 hits in October 2020. The dataset contained two cases twice: Verwaltungsgericht Arnberg [VG] [Administrative Trial Court of Arnberg], Case No. 12 K 768/16.A, (Mar. 10, 2017) (IDs 223 and 226); and Verwaltungsgericht Bayreuth [VG] [Administrative Trial Court of Bayreuth], Case No. B 4 K 16.31692, (Mar. 17, 2017) (IDs 257 and 258). The IDs refer to the internal numbering in the dataset. The decisions with ID 150 and 162 are identical; it was included in the dataset twice by mistake. Only the first ID of these three cases was taken into account. In December 2021, the same query results only in 247 hits. All cases used in this study have been saved by the authors and are available.

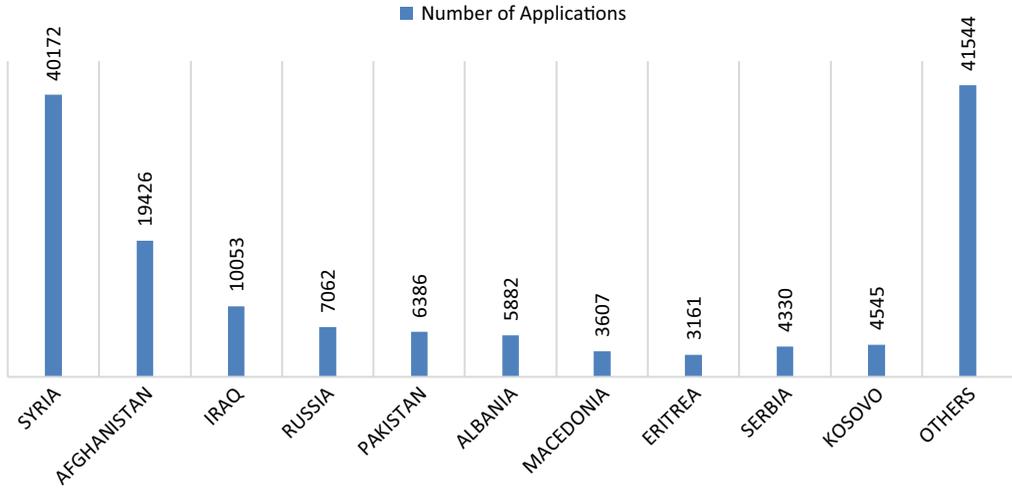


Figure 2. Countries of origin in total in 2017.¹⁰¹

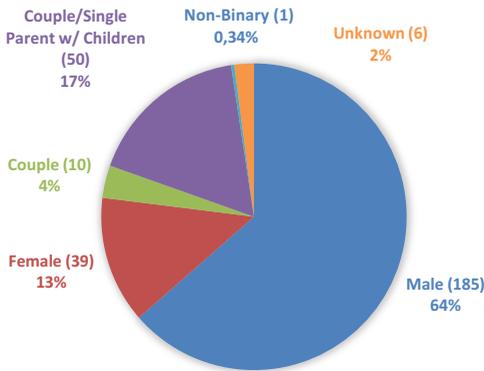


Figure 3. Applicants.

binary, and only ten applications concerned couples without children. In six cases, no such details were available (Fig. 3).

Most decisions in the sample were taken by courts in Bavaria (132) and North Rhine-Westphalia (44). But the sample also includes some decisions from all other German states except for Thuringia (Fig. 4).

3. Limitations

This study has several limitations. *First*, not all the reasons that did in fact have an influence on the court’s credibility assessment must necessarily be put down in writing in the judgment.¹⁰² As the above-mentioned research of *Rehaag* and *Cameron* showed, this might be especially true for reasons that the decision-maker knows to be legally problematic.¹⁰³

¹⁰¹See *Rehaag & Cameron, supra* note 68, at 13; *Millbank, supra* note 27, at 7.

¹⁰²See *Rehaag & Cameron, supra* note 68, at 23.

¹⁰³See *BAFF, supra* note 12, at 19.

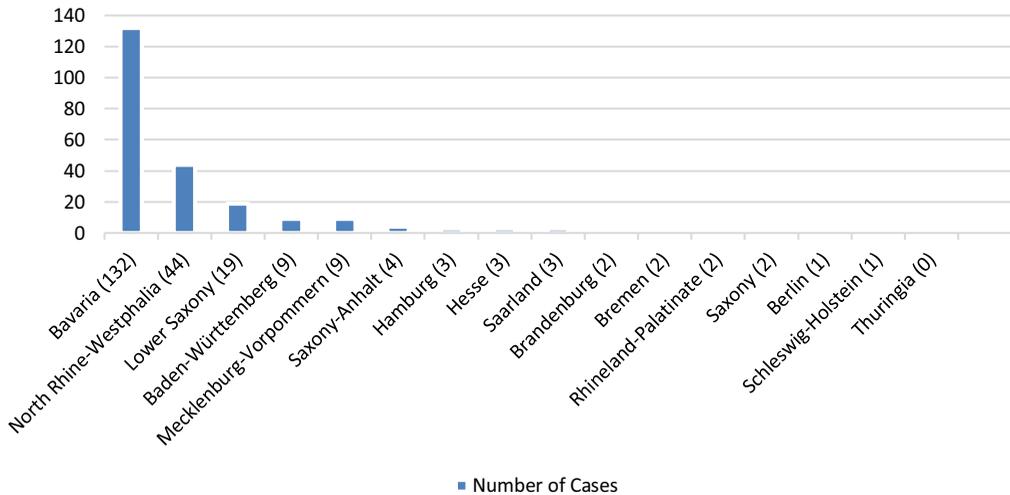


Figure 4. Number of cases decided on by a court in each Land.

Second, a result that was arrived at by other means, for example by intuition, might be rationalized by reference to the credibility criteria *ex post facto*.¹⁰⁴ But this is true for any decision-making procedure that requires the decision-maker to give reasons.

Third, some code values, such as the date of the decision, are rather straightforward. But categorizing the credibility criteria and confounding factors, and the way they were taken into account in a decision, requires judgment. Because courts use various phrases and different wording, depending on the individual case, there was no way to avoid this. It is also the reason why an analysis using corpus linguistics could not be conducted in a sensible manner. The uncertainty introduced by this dependence on judgment, is limited though. In most cases, it is quite clear which credibility criteria the court relied on.

These limitations must be taken into account when analyzing the data and their significance. But it is submitted that valuable insights can be gleaned from the sample analyzed here nonetheless.

II. Results

1. General Overview of the Sample

In 114 of the 236 relevant decisions, the applicant's account was considered credible (48 %). In 105 cases, the account was not considered credible (44 %). Of those 105 cases, the account was considered not credible due to the burden of proof 16 times, meaning the court was not able to convince itself of the credibility to the required degree.

In all 105 cases in which the applicant's account *was not* considered credible, refugee status was denied. In 65 of the 114 cases in which the account *was* considered credible, refugee status was nonetheless denied for other reasons. In 52 of these 65 cases, the courts did not question the account's credibility at all, meaning they did not apply any credibility criteria. Instead, they implicitly found the account credible for the purposes of the decision, and based their legal reasoning on these facts because refugee status had to be denied for other reasons anyway. In another 17 cases, the administrative courts explicitly refrained from deciding on the credibility of the applicant's account because refugee status was rejected for other reasons, most frequently because the

¹⁰⁴See FEDERAL AGENCY, *supra* note 91, at 61.

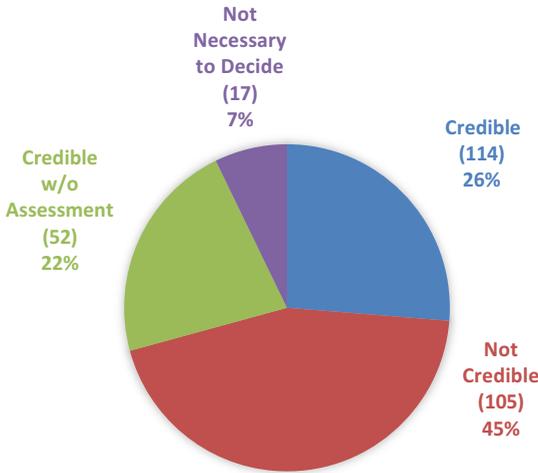


Figure 5. Results of Credibility Assessment.

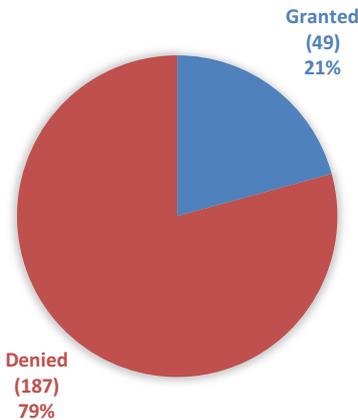


Figure 6. Results of Refugee Status Determination.

account, even if true, did not disclose a persecution ground covered by the Convention or no real risk of persecution.

The sample’s rejection rate (79 %) is close to the refugee status rejection rate of all first-instance court decisions of 2017 with regard to the top ten countries of origin (79.9 %), and lower than the rejection rate of all countries of origin (83.8 %) (Figs. 5 and 6).¹⁰⁵

The grounds for persecution most often argued were religion and political opinion. In most cases, no ground recognized by the Convention was explicitly argued for by the applicant (Table 1).

2. Use of Credibility Criteria

In general, the sample confirms German courts’ reliance on the credibility criteria expounded above. General plausibility, external and internal consistency, the account’s level of detail, and the timeliness of the claim play a dominant role in decision-making. Knowledge is a content-based criterion that seems to be relevant in fewer cases. In addition to these content-related credibility criteria, some courts continue to rely on applicants’ demeanor during the interview and on their

¹⁰⁵See GÖBEL-ZIMMERMANN, EICHHORN & BEICHEL-BENEDETTI, *supra* note 28, at 96; Robert Thomas, *Assessing the Credibility of Asylum Claims: EU and UK Approaches Examined*, 8 EUR. J. OF MIGRATION & L. 79, 81 (2006).

Table 1. Persecution Grounds

	Total Number	Credible	Not Credible	Not Necessary to Decide
Religion	40	16	22	2
Political Opinion	47	35	11	1
Nationality	1	0	1	0
Social Group	13	7	5	1
Social Group (Gender)	2	2	0	0
Social Group (Sexual Orientation)	6	4	2	0
No Convention Ground	97	43	44	10

Table 2. Credibility Criteria Used (Total)

General Plausibility	78
External Consistency	71
Internal Consistency	63
Level of Detail	62
Timeliness of the Claim	47
Demeanor	25
Personal Credibility	16
Knowledge	16
Total:	378

Table 3. External Consistency in More Detail

COI	25
Documents	20
Judicial Perception	9
Media	8
Witnesses	8
Country Expert	1

personal credibility. It has to be noted that one decision can, and often will, rely on multiple criteria (Table 2).

Of the 71 instances in which external consistency was relied on, documents and country of origin information were most often relevant (Table 3).

Decisions that consider the applicant credible refer to many positive credibility criteria. Inversely, decisions that *do not* consider the applicant credible rarely refer to positive credibility criteria, but often to negative credibility criteria.

Most often, an account was considered credible because the courts considered it, in a general sense, plausible (16) or sufficiently detailed (14), because it was supported by documents (14) and/

Table 4. Positive Credibility Criteria

	Applicant Credible	Applicant Not Credible
General Plausibility	16	0
Level of Detail	14	1
Documents	14	1
COI	14	1
Demeanor	13	0
Internal Consistency	10	0
Judicial Perception	8	1
Personal Credibility	6	0
Knowledge	5	1
Timeliness of the Claim	4	0
Media	3	0
Witnesses	3	0
Country Expert	0	1

Table 5. Use of Negative Reality Criteria

	Applicant Credible	Applicant Not Credible
General Plausibility	1	61
Internal Consistency	2	51
Timeliness of the Claim	1	42
Level of Detail	1	46
Demeanor	0	11
COI	0	10
Personal Credibility	0	10
Knowledge	1	9
Media	0	5
Witnesses	0	5
Documents	1	4
Judicial Perception	0	0
Country Expert	0	0

or country-of-origin information (14). The internal consistency of the account was likewise noted quite often (10). The applicant's demeanor during the interview (13) and his or her personal credibility (6) could also have a positive impact. Less often, witness testimony (3), general knowledge (5), and media reports (3) were used. Additionally, the judicial inspection of photos, the applicant's body, and the internet were considered to have no probative force—neither positively nor negatively—in 8 cases (*unergiebig*) (Table 4 and 5).

Table 6. Confounding Factors

	Accepted by the Court	Not Accepted by the Court
Mistake by Interpreter	2	8
Mistake in Interview Transcript	0	4
Other Explanation for Negative Reality Criteria	0	3
Age	0	1
Psychological Trauma	1	1
Cultural Distance	2	1
Gender	0	0
Limited Ability to Express Oneself	2	0
Error	0	0
Total:	7	18

3. Taking into Account Confounding Factors

Confounding factors were comparatively rarely an issue in the sample's decisions. Only 7 times a decision acknowledged that a confounding factor could (partly) counteract a negative credibility criterion. 18 times, the courts rejected a confounding factor that the applicant had advanced. Most often, this was a mistake by the interpreter (Table 6).

4. Refusing to Give (further) Reasons

A tool that was used in 88 cases is the possibility under Section 77(2) Asylum Act for a court to "dispense with a further presentation of the facts and of the reasons for its decision." A court may do so provided it "follows the statements and justification" of the Federal Agency or the parties agreed to it. In 50 of these 88 cases, no convention ground had been argued. In such cases, the courts likely relied on the original decision's facts and reasoning to a greater extent because the result is evident. In 11 of the other 38 cases, the courts accepted applicants' account as credible, relying on Section 77(2) Asylum Act. Finally, in all 24 cases, in which the courts found the applicant's account not credible and relied on Section 77 (2), they did so only partially or additionally. In fact, they did rely on various reality criteria in all of these cases. The courts just agreed with the Federal Agency and stated so by relying on Section 77 (2). Because the reasons given by the Agency are mostly not reiterated in the court's decision, no final assessment of these cases is possible.

III. Analysis

The analysis will evaluate the empirical findings. It will elaborate on their significance for a qualitative analysis of credibility assessment. Doing so, it will also point out, qualitatively, examples from the dataset that illustrate when credibility criteria seem to be applied in a sensible manner—and when they are not, in particular with regard to confounding factors.

1. Confirmation of Content-Based Credibility Criteria

This quantitative analysis confirms that content-based credibility criteria are indeed a defining feature of German jurisprudence in asylum cases. It also confirms that evidence other than

the applicant's testimony is indeed rarely available (1.1.). While inconsistencies and other criteria are important, the often-criticized plausibility criterion dominates German court practice in one form or another (1.2.).

1.1. Generally

The quantitative analysis conducted here confirms the use of the credibility criteria. All of the criteria that the qualitative analysis suggested are highly relevant in practice. Courts refrain from using them only when the case does not hinge on the account's credibility because the result is evident for different reasons. For example, it need not be established whether the account is credible if it does not reveal any convention ground for persecution. Nevertheless, the account's credibility is only a necessary but not a sufficient condition for the application's success: Additionally, the real risk of persecution due to a recognized convention ground must be shown.

Practitioners and scholars generally assume that applications are most often rejected due to internal inconsistencies,¹⁰⁶ not only in Germany.¹⁰⁷ The importance of this criterion was confirmed in this study: 51 cases refer to it in a negative sense. Only general plausibility was referred to more often as a negative criterion.

As a whole, the dataset confirms that, in many cases, the decisive evidence will be the applicant's account.¹⁰⁸ But external consistency was a factor in some cases, in particular documents (20) and country-of-origin information (25) was referred to. Not all sub-categories of external consistency were used often: Medical evidence, for example assessments by medical professionals, was mostly not a relevant factor in the sample. One case featured a privately commissioned (psychological) medical opinion, which was taken into account positively, and in one case a (psychological) medical opinion commissioned by the court was taken into account negatively.¹⁰⁹ The Istanbul Protocol, an international standard for the assessment of torture victims was not relied on in any case. This low relevance of medical evidence in the sample might be owed to the type of cases. It could be, for example, that applicants who dispose of a positive medical opinion according to the Istanbul Protocol are granted refugee status by the Federal Agency as a matter of course and thus need not have the Agency's decision reviewed by a court.

1.2. Dominance of Plausibility

Plausibility criteria dominated the credibility assessment in the sample. The courts referred to plausibility in 203 cases, 53.7 % of all criteria.

a) General Plausibility

In the examined cases, the general plausibility of the account was the criterion most often relied on—78 times in total. The account was considered plausible in 16 cases and not plausible in 62 cases.

General plausibility is often regarded as a problematic credibility criterion because even events that seem improbable or bizarre, may have taken place just like that.¹¹⁰ Four detainees who steal uniforms,

¹⁰⁶See Melanie Griffiths, "Vile Liars and Truth Distorters": Truth, Trust and the Asylum System, 28 ANTHROPOLOGY TODAY 8 (2012) (for Great Britain); GOOD, *supra* note 53, at 187; Catriona Jarvis, *The Judge as Juror Re-visited*, 9 IMMIGR. L. DIG. 16, 21 (2003). See also Carolina Kobelinsky, *Emotions as Evidence*, in OF DOUBT AND PROOF: RITUAL AND LEGAL PRACTICES OF JUDGMENT 163, 170 (Daniela Berti, Anthony Good & Gilles Tarabout eds., 2015) (for France); Refugee Protection Division, ASSESSMENT OF CREDIBILITY IN CLAIMS FOR REFUGEE PROTECTION 23 (2004) (for Canada).

¹⁰⁷See Noll, *supra* note 4, at 607.

¹⁰⁸Verwaltungsgericht München [VG] [Administrative Trial Court of Munich], Case No. M 4 S7 17.33059, (Feb. 27, 2017); Verwaltungsgericht Minden [VG] [Administrative Trial Court of Minden], Case No. 10 K 2488/14.A, (Jan. 12, 2017) (case IDs 186 and 14 respectively).

¹⁰⁹See Millbank, *supra* note 27, at 23; Bingham, *supra* note 62, at 13–14; MARK SYMES & PETER JORRO, ASYLUM LAW AND PRACTICE No. 2.24–2.25 (2nd ed. 2010).

¹¹⁰See DANUTA CZECH, KALENDARIUM DER EREIGNISSE IM KONZENTRATIONSLAGER AUSCHWITZ-BIRKENAU 1939–1945 232 (1989).

weapons, a vehicle, and escape a heavily guarded internment camp might remind the reader of a movie plot. But this is exactly what *Kazimierz Piechowski, Józef Lempart, Stanisław Gustaw Jaster, and Eugeniusz Bendera* did when they fled the concentration camp Auschwitz on June 20, 1942.¹¹¹

Some consider the plausibility criterion “useless” because of the subjective judgment that its application requires.¹¹² It is true that it tends to set the decisionmaker’s experiences as absolute.¹¹³ Theoretically, it allows the decision-maker to speculate how the persecuted person or the persecutors would have acted in a true account.¹¹⁴ Plausibility thus compares the applicant’s account to an ideal—to which reality need not conform.¹¹⁵ Victims of grave human rights violations need not act in a manner that seems rational to the decision-maker in an asylum case.¹¹⁶ This is true, for example, for risky conduct such as the sale of Christian utensils in Afghanistan under the Taliban or a kiss between two men in public where such conduct might incur criminal responsibility.¹¹⁷

Why then, despite these limitations, is the criterion so regularly used in German court practice? The fact that this criterion was used the most in the sample could to some extent be owed to the design of the study: General plausibility served as somewhat of a catch-all criterion for probability judgments that did not fit one of the more specific categories. But there is also another explanation: Plausibility is reminiscent of, and in certain cases consists of, the use of circumstantial evidence. Circumstantial evidence points to a known fact in order to make the existence of another fact more likely, but without making the existence of this other fact a necessity.¹¹⁸ The German Federal Constitutional Court has confirmed that circumstantial evidence may be used in asylum cases as well.¹¹⁹

Consider, for example, the case already mentioned above, in which the applicant claimed that he had wrested himself from a guard, left his clothes in the guard’s hand, then evaded another five of them by vaulting a two-meter wall. These facts beg the question how this was possible:

It takes time to wrest oneself away from a guard so vigorously that clothes are left in his hand and to scale a two-metre wall—yet none of the six soldiers outside intervened to prevent it or to shout for assistance even just from those already outside the house. It was for the Appellant to give such evidence of the disposition of the soldiers, and of the layout of the house, exterior grounds, wall and entrance that might explain how it all might have happened.¹²⁰

While general plausibility has its place in credibility assessment, it is indeed prone to abuse and error. It should be applied restrictively. In order to minimize plausibility’s tendency to invite speculation and prejudice, its application must be based on reliable country-of-origin information.¹²¹

¹¹¹See Noll, *supra* note 4, at 617.

¹¹²See MARTINA RIENZNER, INTERKULTURELLE KOMMUNIKATION IM ASYLVERFAHREN 86 (2011). Cf. Udo Undeutsch, *Glaubhaftigkeit von Aussagen*, in FORENSISCHE PSYCHOLOGIE 134–135 (Udo Undeutsch ed., 1967).

¹¹³See Jarvis, *supra* note 106, at 18. Cf. Thomas, *supra* note 105, at 84; Rehaag & Cameron, *supra* note 68, at 25; Verwaltungsgericht Stuttgart [VG] [Administrative Trial Court of Stuttgart], Case No. A 11 K 2664/10, at 5, (Jan. 24, 2011).

¹¹⁴Cf. HUNGARIAN HELSINKI COMMITTEE, *supra* note 31, at 48.

¹¹⁵HILARY EVANS CAMERON, REFUGEE LAW’S FACT-FINDING CRISIS: TRUTH, RISK, AND THE WRONG MISTAKE 67–68 (2018).

¹¹⁶See *id.* at 121–124 (detailing these examples approved on appeal). For more examples, see Millbank, *supra* note 27, at 20.

¹¹⁷See Regnerus Engelhard, ALLGEMEINES FEINLICHES RECHT 596–602, 643 et seq. (1756) (showcasing that circumstantial evidence has been used in German criminal trials for centuries).

¹¹⁸Bundesverfassungsgericht [BVerfG] [Federal Constitutional Court], 2 BvR 1838/15, para. 33, (Apr. 3, 2020), https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/DE/2020/04/rk20200403_2bvr183815.html (“es sind auch äußere Anknüpfungstatsachen heranzuziehen, die Rückschlüsse auf die innere Einstellung des Betroffenen erlauben”).

¹¹⁹See UK, *MM (DRC – Plausibility)* Democratic Republic of Congo [2005] UKIAT 00019, at 26.

¹²⁰See Hessischer Verwaltungsgerichtshof [VGH] [Higher Administrative Court of Hesse], Case No. 13 UE 3545/89, paras. 44–50, (Mar. 11, 1991) (representing a positive example correcting a negative example by the administration).

¹²¹See Verwaltungsgericht Bremen [VG] [Administrative Trial Court of Bremen], Case No. 1 K 1128/16 (ID 81), para. 22, (Feb. 1, 2017).

b) Timeliness

The claim's timeliness—and the account's level of detail and the applicant's knowledge about certain issues—are specific emanations of the plausibility criterion. Yet, they seem less problematic to many commentators. In a way they are, because their premise, the connection between the fact that is used as circumstantial evidence and the fact that should be proven by it, seem better established. Nonetheless, their application requires great care.

The timeliness of a claim becomes relevant when more acts of persecution or acts of persecution of a qualitatively more serious nature are added to the original account in the course of the administrative procedure or in court. A later submission of persecution-relevant facts requires explanation. The courts referred to the timeliness of the claim 47 times in total. In 43 of those cases, the criterion was deemed negative, meaning that the applicants could not sufficiently explain why they did not present these facts earlier in the proceedings.

In 4 cases, still, the court was convinced by the applicant's explanation. For example, the credibility of an account was not called into question by the fact that the applicant did not mention his activity in the Kurdish regional parliament during his interview at the Federal Agency. The applicant pointed to communication difficulties with the interpreter and that the interpreter had asked him to keep his statements as brief as possible. He also explained that at the time of his hearing, the Federal Office consistently granted refugee status to refugees from Syria. Therefore, he had not insisted on telling his whole story.¹²² In the other cases, the applicant convincingly explained a later claim by pointing out that the Federal Agency had not asked about an issue (religion),¹²³ that the event in question had not happened yet (a Christening),¹²⁴ and that he had assumed that a pattern of persecution (against relatives of defected officers in Syria) was commonly known in the Federal Agency.¹²⁵ It is thus crucial for courts to ask the applicant why a claim is brought forward late.¹²⁶ As the examined cases show, it may well be explainable.

c) General Knowledge

The context-independent general knowledge of the applicant is a credibility criterion if such knowledge can plausibly be expected. This is particularly the case in the assessment of the seriousness of religious conviction and also origin. Such general knowledge was referred to 16 times. In 6 cases, the knowledge of the applicant was deemed positive and in 10 cases the lack of knowledge was deemed negative. For example, one applicant claimed to be part of the ethnic group of Tigrayans in Eritrea, but he did not speak their language, Tigrinya.¹²⁷ In another case, an applicant's religious conversion was considered not credible, despite sufficient knowledge about the Christian religion, just because the court thought the applicant was abusing the asylum system as it held many other Iranian asylum seekers to do.¹²⁸

¹²²See Verwaltungsgericht München [VG] [Administrative Trial Court of Munich], Case No. M 23 K 16.30422 (ID 250), para. 36, (Mar. 16, 2017).

¹²³See Verwaltungsgericht Aachen [VG] [Administrative Trial Court of Aachen], Case No. 7 K 3441/16.A (ID 221), para. 40, (Mar. 10, 2017).

¹²⁴See Verwaltungsgericht Wiesbaden [VG] [Administrative Trial Court of Wiesbaden], Case No. 6 K 2294/16.WI.A (ID 276), para. 18, (Mar. 22, 2017).

¹²⁵See on the duty to confront below.

¹²⁶See Verwaltungsgericht Ansbach [VG] [Administrative Trial Court of Ansbach], Case No. AN 3 S 17.30832 (ID 169), para. 5, (Feb. 22, 2017).

¹²⁷See Verwaltungsgericht Bayreuth [VG] [Administrative Trial Court of Bayreuth], Case No. B 2 K 16.31024 (ID 182), paras. 32, 40, (Feb. 24, 2017).

¹²⁸See Verwaltungsgericht Lüneburg [VG] [Administrative Trial Court of Lueneburg], Case No. 3 A 134/16 (ID 29), para. 20, (Jan. 16, 2017).

d) Level of Detail

The court referred to the level of detail of the applicants' account in 62 cases. In 16 cases, the court positively considered the level of detail and in 47 cases it negatively considered a lack of detail. For example, in one case the applicant could not specify the threats he was said to have received even after he was explicitly asked to do so: "His statements about the threats remained brief, abstract, general, colorless, and superficial. They did not appear to the court to be a description of what he had actually experienced." The court noticed how he was able to describe other situations in more detail.¹²⁹

2. Continued Use of Discredited Conduct-Based Criteria

Some German administrative courts still rely on conduct-based credibility criteria, such as the applicant's demeanor and personal credibility; even though, such criteria have long been discredited in the relevant scientific literature. They should be abandoned entirely.

Personal credibility was in total relied on in 16 cases. But in 6 of the 24 cases in which demeanor was used, personal credibility was assessed as well. So, a decision-maker who relies on demeanor seems more likely to also rely on personal credibility, and vice versa.

This use of demeanor and personal credibility is not confined to individual judges, courts or states. The 24 cases in which demeanor was used stem from 15 different administrative courts in different states. It was used in cases with male, female, and non-binary applicants as well as when couples with children applied for refugee status. The countries of origin concerned were varied and from different continents: Afghanistan, Iran, and Syria, but also Kosovo, Eritrea, and Sri Lanka. While these conduct-based criteria are not (explicitly) used as often as the content-based criteria, they still influenced the courts in the dataset in a considerable number of cases.

2.1. Demeanor

Demeanor was considered 24 times: As a positive factor in 13 cases, and as a negative one in 11. In 16 of these 24 cases, the court referred to the general impression made by the applicant during the interview. In the other cases, emotion (4), aggression (1), and the manner of speech (3) were taken into account.

For example, in one case the applicant's gestures and facial expression were taken into account negatively because he did not react in the way the court expected when he was shown a photo that was supposed to depict his dead relatives: "Appropriate emotions or dismay could neither be seen in the claimant's gestures nor in his facial expression—even after the court told him that these are terrible photos and that the court deplored what could be seen on these photos, if they showed his relatives."¹³⁰ In another case, it was taken into account positively that "[a]s a whole, the claimant made a calm and collected impression. He did not seem to just rehearse knowledge that he had learned by heart. To the various questions [on the Christian faith] he replied equally convincing, without seeking to evoke a certain impression in the court."¹³¹

Researchers have tried to find "cues" in a person's demeanor that indicate a lie for a long time. These attempts have failed. According to the relevant experiments, laypersons using demeanor as an indicator only recognize the truth of a statement with a probability that corresponds to chance.

¹²⁹See Verwaltungsgericht Lüneburg [VG] [Administrative Trial Court of Lueneburg], Case No. 3 A 134/16 (ID 29), para. 26, (Jan. 16, 2017) ("Angemessene Emotionen oder eine Betroffenheit des Klägers war weder in Gestik noch Mimik des Klägers zu erkennen, auch nicht, als das Gericht ihm sagte, dass es schreckliche Fotos seien und es das zu Sehende bedauere, wenn es sich um seine Angehörigen handele.") (translation by the authors).

¹³⁰See Verwaltungsgericht Greifswald [VG] [Administrative Trial Court of Greifswald], Case No. 3 A 374/16 As HGW (ID 40), para. 68, (Jan. 18, 2017) ("Insgesamt wirkte der Kläger während der Befragung ruhig und gefasst. Er erweckte nicht den Eindruck, etwa auswendig gelerntes Wissen abzuspielen. Auf die verschiedenen Fragen reagierte der Kläger gleichermaßen überzeugend, ohne darauf bedacht zu sein, bei Gericht einen bestimmten Eindruck erwecken zu wollen.")

¹³¹See VRIJ, *supra* note 5, at 182–188.

The tossing of a coin would, with equal certainty, distinguish the truth from a lie—regardless of how well one knows the person.¹³² This is widely accepted in the relevant scientific community that conducts empirical studies in this area.¹³³ Some even consider the search for “cues” in people’s demeanor to be “pseudoscience”.¹³⁴ “Lively debates about the merits of nonverbal lie detection no longer take place at the scientific conferences that we attend. Yet nonverbal lie detection remains highly popular among practitioners.”¹³⁵

The main reason why demeanor is no useful indicator seems to be that one’s demeanor during an interview is too individual and can have many different causes.¹³⁶ The formality of the situation can make people insecure and nervous,¹³⁷ in particular if they are vulnerable.¹³⁸ Others may be able to speak more freely and without inhibitions—which may in turn seem suspicious.¹³⁹ Certain “cues” for lying that may seem common sense to laypersons and some decisionmakers, as avoiding eye contact, fidgeting and swallowing due to a dry mouth, have proven to be unreliable as well.¹⁴⁰ Despite many empirical studies, no connection could be established between such demeanor and lies.¹⁴¹ Sometimes studies even lead to contradictory results: Someone who is lying may try to avoid demeanor that is often thought of as a cue, for example the person may keep *more* eye contact than usual. Someone who is not lying also has an interest in being perceived as honest and credible.¹⁴² In particular when it comes to sexual offences the expectation that victims behave in some “typical” fashion has time and again lead to grave mistakes because the one specific demeanor that indicates truthfulness or subterfuge simply does not exist.¹⁴³

This can also be seen in the case that was recounted above: The judge assumed that there is a specific way in which one has to react to pictures of one’s dead relatives. The problem with this is not only that there is no one reaction that a person must have in this situation. Even if the “correct” reaction was had, it might still be taken by the decisionmaker to be *too* emotional or “manipulative,” or as the judge in the above-mentioned case put it “seeking to evoke a certain impression in the court.” Applicants would thus have to show a certain performance that is neither too unemotional but also not too “over the top,” according to the standards of the decision-maker, so that they behave in a way that is perceived as “credible.”

This malpractice can also be observed in cases concerning homosexual applicants. On the one hand, the court may expect a stereotypical appearance and demeanor. On the other hand, the

¹³²See VRIJ, *supra* note 5, at 37–99; Aldert Vrij, Maria Hartwig & Pär Anders Granhag, *Reading Lies: Nonverbal Communication and Deception*, 70 ANN. REV. PSYCH. 295, 301–302 (2019); BENDER, ET AL., *supra* note 7, at 251–252; Revital Ludewig, Daphna Tavor & Sonja Baumer, *Wie können aussagepsychologische Erkenntnisse Richtern, Staatsanwälten und Anwälten helfen?*, AKTUELLE JURISTISCHE PRAXIS [AJP] 1415, 1421–1422 (2011); Lothar Tent, *Psychologische Tatbestandsdiagnostik*, in FORENSISCHE PSYCHOLOGIE 208–209 (Udo Undeutsch ed., 1967); JANSEN, *supra* note 6, at 286 et seq.

¹³³Vincent Denault et al., *The Analysis of Nonverbal Communication: The Dangers of Pseudoscience in Security and Justice Contexts*, ANUARIO DE PSICOLOGIA JURÍDICA 1, 8 (2020) (including statements of fifty-one authors).

¹³⁴See generally Vrij, Hartwig & Granhag, *supra* note 132.

¹³⁵See BENDER, ET AL., *supra* note 7, at 262 et seq.; ERICH DÖHRING, *DIE ERFORSCHUNG DES SACHVERHALTS IM PROZESS: BEWEISERHEBUNG UND BEWEISWÜRDIGUNG* 70–71 (1964).

¹³⁶See Bingham, *supra* note 62, at 10.

¹³⁷See HUBERT HEINOLD, *RECHT FÜR FLÜCHTLINGE* 118 (7th ed. 2015).

¹³⁸See Millbank, *supra* note 27, at 9 (containing information on respective Canadian decisions).

¹³⁹See FRED E. INBAU & JOHN E. REID, *LIE DETECTION AND CRIMINAL INTERROGATION* 154 et seq. (1953) (“symptoms’ of guilt”); FRED E. INBAU, JOHN E. REID, JOSEPH P. BUCKLEY & BRIAN C. JAYNE, *CRIMINAL INTERROGATION AND CONFESSIONS* 121 et seq. (4th ed. 2001).

¹⁴⁰See Vrij, *supra* note 5, at 37–99 (containing further references); Vrij, Hartwig & Granhag, *supra* note 132, at 301 et seq. (comprehensive on the current state of research).

¹⁴¹See BENDER, HÄCKER & SCHWARZ, *supra* note 7, at 358; VRIJ, *supra* note 5, at 53–62, 88–99; JANSEN, *supra* note 6, at 286 et seq.

¹⁴²See VRIJ, *supra* note 5, at 243–244 (with further references).

¹⁴³See Verwaltungsgericht Stuttgart [VG] [Administrative Trial Court of Stuttgart], Case No. A 5 K 3670/16 (ID 271), para. 25, (Mar. 21, 2017) (“tuntenhaftes Auftreten”).

court may consider such appearance and demeanor artificial. For example, in one case the court noted positively that the applicant did *not* try to appear as stereotypically homosexual.¹⁴⁴

Expecting such performances evidently does not serve the purpose of credibility determination. To the contrary, as the British judge *Thomas Bingham* stated long ago: “To rely on demeanour is in most cases to attach importance to deviations from a norm when there is in truth no norm.”¹⁴⁵ For asylum cases, in which cultural distance between the decision-maker and the applicant exists, and language interpretation is most often necessary, this is all the more true.

Why then is demeanor still used? A major reason seems to be a lack of interdisciplinary work. The results of psychological research in this area simply have not been taken into account adequately in legal literature and practice. As the law stands, demeanor may be taken into account as a credibility criterion. Courts and legislators accept it.¹⁴⁶

The law and the court practice must be changed to reflect the insight that demeanor is an invalid criterion that cannot be used to divide credible from incredible accounts. The Federal Agency and administrative courts must not rely on it anymore.

2.2. Personal Credibility

The idea that certain “types” of witnesses exist, that their station in society or their morals are relevant for assessing the credibility of their account is old.¹⁴⁷ Most commentators reject the idea now.¹⁴⁸ Yet, the idea that, for other reasons, the applicant in an asylum case could be considered personally credible or incredible, survives till this day in German court practice, even though the threshold for considering an applicant personally incredible is held to be high.¹⁴⁹

In 16 cases, the court referred to the personal credibility of the applicant. In 10 of these cases, the court deemed the claimant not credible, whereas in 6 cases the person was held to be credible. The court mostly just postulated the personal credibility without giving an explanation. If reasons were given, personal credibility was determined based on the content-based criteria.

As intuitive as this criterion of personal or general credibility of a person may seem,¹⁵⁰ it does not have analytical value beyond what can already be taken into account through the content-based criteria. Even worse, it provokes mistakes. This is most obviously true for stereotypical prejudice that seeks to draw conclusions from a person’s looks or “type.”¹⁵¹ In particular in asylum claims, personal credibility is extremely prone to reflect such prejudice.¹⁵²

Some people may indeed be more likely than others to say the truth in a certain situation. But no one is constitutively credible or incredible.¹⁵³ Extrapolating from a person’s personal credibility to the credibility of his or her specific account is always a mistake.¹⁵⁴ Even a person that has been

¹⁴⁴See *Bingham*, *supra* note 62, at 11. See also *SYMES & JORRO*, *supra* note 109, at No. 2.37. Cf. *Kälin*, *supra* note 59, at 231–232.

¹⁴⁵See *supra* sect. B.I.2.

¹⁴⁶See CARL JOSEPH ANTON MITTERMAIER, *DIE LEHRE VOM BEWEISE IM DEUTSCHEN STRAFPROZESSE* 344–345 (1834).

¹⁴⁷See *BENDER*, *supra* note 19, at 84–85; *Noll*, *supra* note 4, at 614.

¹⁴⁸See Bundesverfassungsgericht [BVerfG] [Federal Constitutional Court], Jan. 9, 1963, 15 *ENTSCHEIDUNGEN DES BUNDESVERFASSUNGSGERICHTS* [BVerfGE] 249, 254; *GUSY*, *supra* note 69, at 60.

¹⁴⁹See *GÖBEL-ZIMMERMANN, ET AL.*, *supra* note 28, at 148 (“persönlicher Eindruck”); *ROMAN FRÄNKEL, AUSLÄNDERRECHT* § 24 AsylG, para. 23 (Rainer M. Hofmann ed., 2nd ed. 2016). See also *HEINRICH LANG, VERWALTUNGSGERICHTSORDNUNG GROSSKOMMENTAR* § 96, para. 26 (Helge Sodan & Jan Ziekow eds., 5th ed. 2018).

¹⁵⁰See *KARL PETERS, STRAFPROZESS* 389 (4th ed. 1985).

¹⁵¹See *BAfF*, *supra* note 12, at 57.

¹⁵²See *Undeutsch*, *supra* note 112, at 52 et seq.; *BENDER, ET AL.*, *supra* note 7, at 254 et seq.; *BAfF*, *supra* note 12, at 57; see also *DÖHRING*, *supra* note 135, at 136.

¹⁵³Cf. *Engelhard*, *supra* note 117, at 619–620 (rejecting already in 1756 the idea that someone’s ‘bad way of life’ is circumstantial evidence against that person).

¹⁵⁴See *R v. Lucas* [1981] 2 All ER 1008 (Eng.) (highlighting the “Lucas direction” in British criminal law that instructs juries to do just that).

caught in a lie is not per se incredible. One lie does not render the entire account incredible.¹⁵⁵ This is generally accepted by national agencies and courts,¹⁵⁶ and also by the UNHCR.¹⁵⁷ Someone who is truthful regarding the core of the account might still lie in other areas, e.g., to meet presumed expectations of the interviewer, out of shame, or in order to unnecessarily improve their position in the asylum proceedings.¹⁵⁸

Lies can be taken into account as inconsistencies in the content-based assessment of the account's credibility. But their significance must be weighed in the individual case, taking into account confounding factors.¹⁵⁹ A category of personal credibility that focuses on the person of the applicant is superfluous in so far as it relates to aspects that can be taken into account with content-based criteria. Beyond that it is prone to mistakes and abuse. It should therefore be rejected as a criterion of credibility assessment *de lege ferenda*.¹⁶⁰

3. Correlation of Credibility Criteria and Credibility

The assessment of some credibility criteria as positive or negative correlates strongly with the court's conviction concerning the applicant's credibility. *First*, as already mentioned, in all cases the positive or negative assessment of conduct-based criteria corresponds to the court's conviction. *Second*, the number of cases in which the plausibility criteria were deemed positive is almost the same as the number of cases in which the court considered the applicant credible and vice versa. This could suggest two things: That a result that was arrived at by other means, such as simple prejudice, was rationalized by reference to these criteria *ex post facto*. Or these criteria were used to overcome remaining doubts in one or the other direction.

4. Insufficient Awareness of Confounding Factors and the Duty to Confront

Negative credibility criteria indicate an incredible account only in a *prima facie* manner. The account may display negative credibility criteria despite being truthful. There might be other explanations for contradictions and implausibilities. It is therefore always necessary to consider the possibility that the applicant's account exhibits negative credibility criteria only due to confounding factors.¹⁶¹ Such factors were insufficiently taken into account in the court practice reflected in the dataset. While this study of course could not reveal whether confounding factors were accurately accepted or rejected, it stands to reason that the possibility of confounding factors should have been at least considered in many more than the 25 cases that did so.

¹⁵⁵See, e.g., Hessischer Verwaltungsgerichtshof [Higher Administrative Court of Hesse], Case No. 13 UE 3545/89, para. 51, (Mar. 25, 1994); Verwaltungsgericht Frankfurt [VG] [Administrative Trial Court of Frankfurt], Case No. 3 E 30495/98.A (2), paras. 24–26 (Aug. 29, 2001); UK Home Office, *supra* note 45, at 12–13; IARLJ, *supra* note 15, at 77; Turcios v. INS, 821 F.2d 1396 (9th Cir. 1987); STEPHEN H. LEGOMSKY & DAVID B. THRONSON, IMMIGRATION AND REFUGEE LAW AND POLICY 1288 (7th ed. 2019).

¹⁵⁶See United Nations High Comm'r for Refugees, Handbook on Proc. and Criteria for Determining Refugee Status and Guidelines on Int'l Prot. under the 1951 Convention and the 1967 Protocol Relating to the Status of Refugees (2019), para. 199.

¹⁵⁷See SONJA PÖLLBAUER, "I DON'T UNDERSTAND YOUR ENGLISH, MISS." DOLMETSCHEN BEI ASYLANHÖRUNGEN 119–220 (2005); Jarvis, *supra* note 113, at 17; Refugee Protection Division, *supra* note 106, at 11. See also CAMERON, *supra* note 115, at 61.

¹⁵⁸See MICHAEL FUNKE-KAISER, GEMEINSCHAFTSKOMMENTAR ZUM ASYLGESETZ [GK-AsylG] AsylVfG 1992, Vor II -3 GK-para. 229 (2003).

¹⁵⁹See Verwaltungsgerichtshof Baden-Württemberg [Higher Administrative Court of Baden-Wuerttemberg], Case No. A 11 S 2526/17, para. 7, (Nov. 22, 2017) (also rejecting). See also INFORMATIONSVERBUND ASYL UND MIGRATION, ANHÖRUNG IM ASYLVERFAHREN 7 (4th ed. 2016) (noting that in the voluntary counselling practice, there is a fear that, contrary to what is stated here, individual lies in the asylum procedure could be taken as a direct reason to find the entire claim to be implausible); MARX, *supra* note 38, at § 24 para. 21 (same).

¹⁶⁰See *supra* sect. B.II. and III.

¹⁶¹See Verwaltungsgericht Regensburg [VG] [Administrative Trial Court of Regensburg], Case No. RO 12 K 16.32240 (ID 11), para. 6, (Jan. 10, 2017).

Legally, courts have a duty to confront applicants with negative credibility criteria, in order to rule out that they are based on confounding factors. In one case, the court noted the applicant's claim that the interpreter in the administrative procedure did not properly speak the applicant's language, but ultimately the court simply ignored it.¹⁶² Confounding factors must be considered by decision-makers not only when applicant or counsel argue for it. They must be considered *proprio motu* whenever there is sufficient reason to do so.¹⁶³ Article 4 (1) QD requires EU Member States to cooperate with applicants to assess the relevant elements of the application:

[If] for any reason whatsoever, the elements provided by an applicant for international protection are not complete, up to date or relevant, it is necessary for the Member State concerned to cooperate actively with the applicant, at that stage of the procedure, so that all the elements needed to substantiate the application may be assembled.¹⁶⁴

According to Article 16 Asylum Procedures Directive (APD),¹⁶⁵ the applicant must not only be "given an adequate opportunity to present elements needed to substantiate the application" but also "the opportunity to give an explanation regarding elements which may be missing and/or any inconsistencies or contradictions in the applicant's statements." So, for example, an application may be rejected for a lack of details only if the applicant was asked for them.¹⁶⁶ For evident contradictions, the significance of which could not possibly escape the applicant, an exception might be accepted.

Yet, confounding factors were rather rarely taken into account in the sample. Of all 239 cases, only 7 times the existence of a confounding factor was acknowledged. 18 times, the courts rejected the possibility that a confounding factor could have had an influence. It stands to reason that the courts should have considered the possibility of confounding factors in many more cases. Decision-makers apparently do not take their duty to confront the applicant with negative credibility criteria, emanating from Article 4 (1) QD and Article 16 APD,¹⁶⁷ sufficiently seriously. They need to consider confounding factors, if there is any reason to do so. They must point out negative credibility criteria to allow for confounding factors to become apparent.

The cases in which the courts did consider confounding factors show how important it is to do so. But these cases also show that considering confounding factors need not unduly complicate matters by giving someone caught in a lie an additional line of defense, for example the opportunity to advance potentially far-fetched, self-serving justifications. Such justifications—or "(self-)protective assertions," *Schutzbehauptungen*—can of course be rejected, according to the same credibility assessment.

For example, in one case, the applicant argued that he did not disclose earlier that his brother was actively opposing the Syrian regime because he was afraid that the information would be relayed to his home country's authorities by the Federal Agency. The court acknowledged that he may have been excited and even confused during his interview with the Agency. But the court considered the applicant's justification "absurd" that he had not revealed this information because he thought that the Federal Agency had connections to the Syrian regime and might harm his family. The court pointed to the interview's protocol which stated that he had been asked repeatedly to disclose any reason why he would be in danger upon his return.¹⁶⁸ This qualification as

¹⁶²See IARLJ, *supra* note 15, at 35–36; HUNGARIAN HELSINKI COMMITTEE, *supra* note 31, at 40.

¹⁶³CJEU, Case C-277/11, *M.M. v. Minister of Justice, Equality and Law Reform et al.*, ECLI:EU:C:2012:744 (Nov. 22, 2012), para. 65 f, <https://curia.europa.eu/juris/liste.jsf?num=C-277/11&language=EN>.

¹⁶⁴Directive 2013/32/EU of the European Parliament and of the Council of 26 June 2013 on Common Procedures for Granting and Withdrawing International Protection, 2013 O.J. (L 180), 60–61.

¹⁶⁵See HEINOLD, *supra* note 137, at 119.

¹⁶⁶See *id.*

¹⁶⁷See Verwaltungsgericht Dresden [VG] [Administrative Trial Court of Dresden], Case No. 4 K 689/16.A (ID 188), paras. 10, 100, (Mar. 1, 2017) ("abwegig").

¹⁶⁸See Verwaltungsgericht Lüneburg [VG] [Administrative Trial Court of Lueneburg], Case No. 3 A 194/16 (ID 30), paras. 21–24, (Jan. 16, 2017).

“absurd” is acceptable only if it is taken to mean that the applicant could not have possibly believed that. After all, for its qualification as a confounding factor, it would be irrelevant whether the applicant’s honestly held suspicion would be “absurd” or not. But this reaction of the court also shows a certain tendency to dismiss confounding factors maybe too easily. They should be rejected only after careful consideration.

A good example of how confounding factors should be considered is the following case from the sample. The Afghan applicant argued that he had been attacked by the Taliban. A lack of internal consistency and a low level of detail weighed against his account’s credibility. In the interview at the Federal Agency, he had said that he had been stopped twice by the Taliban. He had been threatened the first time and hurt with a knife the second time. Before the court he later mentioned—even when specifically asked—only one incident and ultimately, reconsidering during the interview, no knife attack. He also made different statements on whether he was alone and if the Taliban were on motorbikes or standing on the street. His account lacked detail. He could not explain why he had stopped at all. Even when told that the court needed details, he only said that he “had been hit” by the Taliban. When asked by his legal counsel, he could not substantiate his account, and merely chose from options that his counsel proposed to him through his questions. The applicant drew a sketch of the situation only more or less against his will and with a lot of help by his counsel. The applicant was more detailed when describing the weapons, but the description did not connect to the specific incident and the court assumed that he had seen such weaponry on other occasions.¹⁶⁹

Finally, and importantly, the court considered whether this inconsistent account that lacked detail was owed to the “personality” of the applicant, for example, how he generally speaks of the things he has experienced. But this possibility was rejected because the applicant was able to speak in detail and vividly about situations that he had experienced: About his work at a bank and also how he went to the market twice a week. The court noted how his legal counsel even had to stop him from continuing on about these things. Because of the inconsistency and the lack of detail of the account relating to his alleged persecution, the court considered the account not credible.¹⁷⁰ This case is a good example of how credibility assessment should be conducted: Carefully weighing the different credibility criteria and inquiring into confounding factors.

5. Bias

Female applicants and couples with children were somewhat more likely to be believed than men. But the small sample size—all the more for couples and non-binary applicants—and the applicants’ different countries of origin, which have different acceptance rates, make it impossible to know if this is significant. A future study would need a larger sample size and would need to focus on specific countries of origin and time frames, maybe even specific persecution grounds, so as to make the circumstances of the cases as comparable as possible.

It could also be considered to look for racist bias according to the different countries of origin, which may indicate—to a certain degree—how applicants would be perceived or “racialized” by decision-makers. It seems difficult, however, to compare the different countries of origin in this regard. Many statistical confounding factors come into play, inter alia, the different circumstances in the countries of origin, different grounds for persecution, etc. (Table 7).

6. Insignificance of the Standard and Burden of Proof and of Article 4 (5) QD in Practice

Two issues that are often and intensely discussed in the literature did not have any significance in the sample: The standard and burden of proof, and Article 4 (5) QD. In 16 cases, courts formally

¹⁶⁹See *id.* at para 25.

¹⁷⁰See, e.g., Berlit, et al., *supra* note 24, at 652–653; IARLJ, *supra* note 15, at 48–49; UNHCR, *supra* note 156, at 196, 203–204; HUNGARIAN HELSINKI COMMITTEE, *supra* note 31, at 46.

Table 7. Bias

	Male	Female	Couple	Couple/Single Parent w/ Children	Non-Binary
Credible	67	17	2	27	1
Not Credible	76	11	5	13	0
Not Necessary to Decide	12	1	2	2	
Credible in % of Cases	43,23 %	58,62 %	22,22 %	64,29 %	100 %

relied on the burden of proof to reject the application, meaning the court was not able to convince itself of the credibility to the required degree, but was also not convinced that the applicant was not credible. But these 16 cases all concern so-called safe countries of origin. These are countries of origin that the German Parliament has designated as generally safe according to Article 16 (3) of the Basic Law. According to Section 29a (2) Asylum Act, in these cases, applicants have the burden to show that they are individually persecuted despite the presumption against persecution based on the generally safe situation in this country. In all other cases, the courts never relied on the burden of proof and the fact that the applicant had failed to meet the standard of proof. Rather, the courts either held the applicant's account to be credible or explicitly rejected the account's credibility.

Article 4 (5) QD is much discussed in literature as a crucial albeit flawed evidentiary rule.¹⁷¹ When it is the applicant's duty to substantiate the application, certain aspects of the applicant's statements shall not need confirmation even if they are not supported by documentary or other evidence. But this rule played no role whatsoever in any of the 236 cases. The reason for this lies in the jurisprudence of the German Federal Administrative Court which unconditionally grants what Article 4 (5) QD promises only in case five criteria have been complied with. While the German courts need to be "fully convinced" of the existence of the relevant facts according to Section 108 (1) VwGO (*volle Überzeugungsgewissheit*), they cannot require something impossible from applicants. Because asylum seekers will typically lack evidence to corroborate their accounts (*sachty-pische Beweisnot*), the Federal Administrative Court holds that an applicant's account must suffice to convince the court if it is credible. Thus, German courts must not reject a claim because the applicant's account was not backed by further evidence anyway,¹⁷² and Article 4 (5) QD serves no purposes in this context.

7. Significance of Procedural Safeguards

The data seems to indicate the necessity to provide applicants with legal counsel and to make more use of the limited possibility to appeal court decisions.

7.1. Legal Counsel

The available data, unfortunately, do not reveal if the applicants were represented by counsel. But there is one observation that makes it likely that many of them were not. In 97 cases, no convention ground of persecution was argued, and refugee status rejected. Despite the fact, that the application could have been rejected on that account alone, in 43 of those cases the applicant's account was considered credible, and in 44 cases not credible. In only 10 of these cases, the courts did not find it necessary to decide on credibility. The reason for this may be that often the courts will try to base their decision on as many grounds as possible so as to be sure that the result is correct and also to insulate it from appeal. But

¹⁷¹See Bundesverwaltungsgericht [BVerwG] [Federal Administrative Court], Apr. 16, 1985, 71 ENTSCHEIDUNGEN DES BUNDESVERWALTUNGSGERICHTS [BVerwGE] 180–183, para. 16. See also GÖBEL-ZIMMERMANN, EICHHORN & BEICHEL-BENEDETTI, *supra* note 28, at 678.

¹⁷²See Section 78 Asylum Act.

unlike in other areas of law, the possibility of an appeal will usually not be a prominent aspect in lower courts' thinking: Appeals in asylum cases are severely restricted in Germany.¹⁷³

The fact that in 97 of all 236 cases the applicants did not argue a convention ground but claimed refugee status nonetheless seems to indicate a lack of legal counselling. While it is possible that in some cases, there may be legal uncertainty if persecution is based on a convention ground, in most cases it will be quite clear if someone argues that they are persecuted for political, religious, or other reasons covered by the Geneva Convention. It can be assumed that legal counsel would not recommend claiming refugee status without indicating a recognized convention ground for persecution. A possible alternative explanation would be that the applicants were simply being honest—even if that meant that their application would certainly be rejected.

7.2. Appeal

In only 5 out of all 236 cases, a judgment was rendered on appeal.¹⁷⁴ All of those appeals were dismissed by the higher administrative courts, once by the Federal Administrative Court.¹⁷⁵ In 4 cases, the credibility assessment of the lower administrative court was not dealt with on appeal.¹⁷⁶ In one case, the appellant claimed a denial of the right to be heard due to a so-called surprise decision (*Überraschungsentscheidung*), but not with regard to the credibility assessment. The higher administrative court nonetheless stated, that "it is self-evident that the person's credibility and the account's credibility are always at stake in asylum proceedings, insofar as they are relevant to the decision."¹⁷⁷

It was to be expected that the number of judgments reviewed in an appeal procedure would be small, because access to judicial remedies beyond the first instance has been severely restricted in German asylum law. Thus, errors that occur in the first instance credibility assessment can only be checked and—if necessary—corrected to a very limited extent. So, the final decision about a person's and their account's credibility often lies in the hands of first-instance individual judges.

Yet, some judgments show the need for a review by a higher administrative court. For example, in one case, the court was not convinced that the applicant had genuinely converted to Christianity because the applicant had not sufficiently reflected on the rape for which he had been convicted and had not apologized to the victim.¹⁷⁸ The court thus disregarded the generally accepted credibility criteria for ascertaining whether someone genuinely converted to a new religion. Instead, the court assessed the applicant's credibility based on what the court considers to be proper conduct for a "good Christian." Of course, this is not relevant in any way for assessing the applicant's credibility. This judgment was in dire need to be reviewed for arbitrariness. While an appeal would not have been possible, an arbitrary credibility assessment can be challenged with a complaint to the Federal Constitutional Court¹⁷⁹ and, if need be, to the European Court of Human Rights.¹⁸⁰

¹⁷³These judgments do not form a part of the original sample.

¹⁷⁴See Verwaltungsgeschichtshof München [VGH] [Higher Administrative Court of Munich], Case No. 20 ZB 17.30396, (Apr. 11, 2017); Oberverwaltungsgericht Nordrhein-Westfalen [OVG] [Higher Administrative Court of Northrhine-Westphalia], Case No. 13 A 923/17.A, (Sep. 5, 2017); Oberverwaltungsgericht Nordrhein-Westfalen [OVG] [Higher Administrative Court of Northrhine-Westphalia], Case No. 19 A 552/17.A (Mar. 29, 2018); Verwaltungsgeschichtshof München [VGH] [Higher Administrative Court of Munich], Case No. 10 ZB 17.30437, para. 3, (Feb. 15, 2018); Bundesverwaltungsgericht [BVerwG] [Federal Administrative Court], Case No. 1 C 33/18, (July 4, 2019).

¹⁷⁵See VGH München, Case No. 20 ZB 17.30396; Bundesverwaltungsgericht [BVerwG] [Federal Administrative Court], Case No. 1 C 33/18; OVG Nordrhein-Westfalen, Case No. 13 A 923/17.A; OVG Nordrhein-Westfalen, Case No. 19 A 552/17.A.

¹⁷⁶See VGH München, Case No. 10 ZB 17.30437, para. 3 (translation by the authors).

¹⁷⁷See Verwaltungsgericht Arnberg [VG] [Administrative Trial Court of Arnberg], Case No. AN 1 K 16.30555 (ID 146), paras. 89 et seq., (Feb. 17, 2017).

¹⁷⁸See, e.g., Bundesverfassungsgericht [BVerfG] [Federal Constitutional Court], Case No. 2 BvR 939/20, para. 22, (July 23, 2020), https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/DE/2020/07/rk20200723_2bvr093920.html; Bundesverfassungsgericht [BVerfG] [Federal Constitutional Court], Case No. 2 BvR 78/08, (Apr. 29, 2009), https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/DE/2009/04/rk20090429_2bvr007808.html.

¹⁷⁹See *R.C. v. Sweden*, App. No. 41827/07, paras. 52–53.

¹⁸⁰See Noll, *supra* note 4, at 620–622.

E. Concluding Observations on the Limits of Objectivity in Decision-Makers and Algorithms

The Oxford Handbook of Refugee Law of 2021 came to the conclusion that: “Evidential assessment in the asylum procedure is dysfunctional.”¹⁸¹ The study conducted here confirms that substantial problems persist. But it also shows how German courts try to assess applicants’ credibility in a manner that is as rational and objective as possible. The aim should be to further standardize, refine, and rationalize the procedure to safeguard it against errors and abuse. The most important step in this regard would be to abolish any reliance on conduct-based criteria.

As any balancing exercise, the application of credibility criteria and confounding factors requires judgment. This judgment is necessarily subjective, which makes it important who takes the decision.¹⁸² Some consider this character of credibility assessment to be incompatible with the rule of law: “In the asylum field, it is not law that rules, but individual decision-makers.”¹⁸³ The criteria are said to be merely a “door-opener” for subjective discretion. But the same critique would be true for any credibility assessment that is not entirely mechanistic. In human rights law, for example, the balancing required by the proportionality analysis is likewise often criticized as highly subjective. Nonetheless, it is an important part of that law.

Credibility assessment is a highly complex task. No reasonable method of credibility assessment that sufficiently takes into account this complexity could eliminate the necessity for epistemic judgment and thus subjectivity. Subjectivity should not be equated with arbitrariness though. It cannot be a requirement of the rule of law that all decision-makers would in any one case come to the exact same conclusion. This asks something of legal decision-making that no system anywhere has ever or could ever achieve.

In fact, aiming to eliminate from credibility assessment the subjectivity of epistemic judgment may lead to procedures that claim to be more “objective” by virtue of requiring less subjective judgment in the individual case. While some of these efforts may reasonably contribute to credibility assessment, time and again efforts have been made in this direction that are not only problematic from a human rights point of view. They lead to a peculiar form of decision-making which portrays itself as objective but is in fact arbitrary.

In particular, technical means are often sought to make credibility assessment more objective. For example, authorities in the Czech Republic used a “phallometric” procedure, or genital plethysmography, to determine the sexual orientation of, mostly male, applicants.¹⁸⁴ While the blood circulation in the applicants’ genitalia was measured, they were shown pornographic material. UNHCR rejected this practice as a violation of fundamental human rights.¹⁸⁵ The Court of Justice of the European Union ruled in *A., B. and C.* that such “tests” violate the right to human dignity enshrined in Article 1 of the EU Charter of Fundamental Rights.¹⁸⁶ General Advocate *Eleanor Sharpston* had rightly pointed out that this was a “particularly dubious” and “pseudo-medical” test.¹⁸⁷

While it may seem obvious that phallometry could not in any way contribute to a credibility assessment, it is but one symptom of the desire to make credibility assessment as objective as possible. Unlike the content-based credibility assessment analyzed here, phallometry seems to

¹⁸¹See Jaya Ramji-Nogales, Andrew I. Schoenholtz & Philip G. Schrag, *Refugee Roulette: Disparities in Asylum Adjudication*, 60 STAN. L. REV. 295, 296 (2007).

¹⁸²See Noll, *supra* note 4, at 619–620.

¹⁸³See generally UNHCR, COMMENTS ON THE PRACTICE OF PHALLOMETRY IN THE CZECH REPUBLIC TO DETERMINE THE CREDIBILITY OF ASYLUM CLAIMS BASED ON PERSECUTION DUE TO SEXUAL ORIENTATION (2011).

¹⁸⁴See UNHCR, *supra* note 29, at 65.

¹⁸⁵ECJ, *A, B and C*, Cases C-148/13 and C-150/13 para. 65.

¹⁸⁶Opinion of Advocate General Sharpston, *supra* note 23, para. 62.

¹⁸⁷See Ann-Kathrin Nezik, iBorderCtrl: *Falsch geblinzelt*, DIE ZEIT (Aug. 27, 2020), <https://www.zeit.de/2020/36/iborder-ctrl-luegen-detektor-software-algorithmus-forschung-kritik>.

require basically no human judgment. Objective measurements seemingly translate into objective decisions, untainted by any subjective assessments.

With the same aim of objectivity, the EU has funded the research project “iBorderCtrl” by European Dynamics with 4.5 million Euros.¹⁸⁸ An Automatic Deception Detection System (ADDS) was supposed to be developed: “ADDS quantifies the probability of deceit in interviews by analysing interviewees’ non-verbal micro-gestures.”¹⁸⁹ iBorderCtrl has reportedly been tested, on a voluntary basis, at European borders.¹⁹⁰ The draft of a Regulation on Artificial Intelligence of 2021 qualifies the use of artificial intelligence in asylum law as a high risk technology, but does not rule out the use of “lie detectors.”¹⁹¹ Rather, it points to that possibility.

As we have seen above, the reason why the conduct of the applicant during an interview was rejected as a credibility criterion was not that decision-makers could not perceive precisely enough the applicant’s conduct. Machines might indeed do so more accurately. The reason was that this conduct is too individual and ambivalent to draw from it any conclusion regarding the account’s credibility. There is absolutely no reason why algorithms should achieve better results here. Rather, algorithms would again be trained to recognize some kind of norm which the individual applicant need not correspond to. Even worse, the results of such technology are often regarded as particularly objective—because no human judgment was involved in its application. A credibility criterion that has been resoundingly rejected as unreliable by the relevant scientific community would thus not only be resurrected but given the technical veneer of particular objectivity.

The credibility assessment conducted by human decision-makers in asylum cases is far from perfect. It can be misused and abused. Nonetheless, the application of content-based credibility criteria can also produce convincing results, as some of the cases described above show. Decisions on credibility cannot be taken without subjective judgment. While human decision-making will never be entirely free from bias, decisions should nonetheless be understood to be objective in a meaningful sense if they rely on established credibility criteria, take into account confounding factors, and do not resort to discredited criteria or prejudice.

Finally, credibility assessment cannot be required to dispel all doubts. This would ask too much of applicants and decision-makers alike. Doubts that remain after careful balancing of credibility criteria and confounding factors are, as the UNHCR, the ECtHR and others emphasize, to be resolved to the applicant’s benefit.

Acknowledgements. The authors would like to thank Dr. Konstantin Chatziathanasiou for his valuable comments on an earlier draft of this article.

¹⁸⁸iBorderCtrl, Technical Framework, <https://web.archive.org/web/20211203234007/https://www.iborderctrl.eu/Technical-Framework>; See European Commission, Intelligent Portable Boreder Control System, <https://cordis.europa.eu/project/id/700626/en> (on EU funding for this project).

¹⁸⁹See Matthias Kolb, *Künstliche Intelligenz: EU testet Lügendetektor an den Grenzen*, SÜDDEUTSCHE (Aug. 5. 2018), <https://www.sueddeutsche.de/digital/grenze-kuenstliche-intelligenz-software-iborderctrl-1.4196243>.

¹⁹⁰*Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, at Recital 39, Annex III, para. 7, COM(2021) 206 final (Apr. 21, 2021) (“Migration, asylum and border control management: (a) AI systems intended to be used by competent public authorities as polygraphs and similar tools or to detect the emotional state of a natural person ...”).

¹⁹¹FEDERAL AGENCY, DAS BUNDESAMT IN ZAHLEN 2017: ASYL, MIGRATION UND INTEGRATION (Nuremberg 2018) 61.