PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES

Improving forensic perpetrator identification with Super-Recognizers

Maren Mayer[a] and Meike Ramon[b,1]

Edited by Thomas Albright, Salk Institute for Biological Studies, La Jolla, CA; received December 3, 2022; accepted March 22, 2023

About a decade ago, Super-Recognizers (SRs) were first described as individuals with exceptional face identity processing abilities. Since then, various tests have been developed or adapted to assess individuals' abilities and identify SRs. The extant literature suggests that SRs may be beneficial in police tasks requiring individual identification. However, in reality, the performance of SRs has never been examined using authentic forensic material. This not only limits the external validity of test procedures used to identify SRs, but also claims concerning their deployment in policing. Here, we report the first-ever investigation of SRs' ability to identify perpetrators using authentic case material. We report the data of 73 SRs and 45 control participants. These include (a) performance on three challenging tests of face identity processing recommended by Ramon (2021) for SR identification; (b) performance for perpetrator identification using four CCTV sequences depicting five perpetrators and police line-ups created for criminal investigation purposes. Our findings demonstrate that the face identity processing tests used here are valid in measuring such abilities and identifying SRs. Moreover, SRs excel at perpetrator identification relative to control participants, with more correct perpetrator identifications, the better their performance across lab tests. These results provide external validity for the recently proposed diagnostic framework and its tests used for SR identification (Ramon, 2021). This study provides the first empirical evidence that SRs identified using these measures can be beneficial for forensic perpetrator identification. We discuss theoretical and practical implications for law enforcement, whose procedures can be improved via a human-centric approach centered around individuals with superior abilities.

perpetrator identification | face identity processing | criminal investigation | CCTV crime footage | forensic face matching

Significance

Forensic perpetrator identification is a crucial process depending on accurate processing of the available evidence. In many cases where perpetrators are captured via CCTV footage of a crime, this material is challenging to process—typically due to low quality and limited information available. Over the past decade, the media has suggested that so-called Super-Recognizers (SRs), individuals with superior ability for processing facial identity, could improve perpetrator identification for law enforcement. In reality, such reports are anecdotal success stories—a formal test of SRs' ability has never been conducted. Using authentic forensic case material, we performed the first-ever comparison between SRs and control participants to answer the question: Can SRs improve perpetrator identification accuracy?

Super-Recognizers (SRs) were first described by Russell et al. (1) as individuals with exceptional abilities in face perception and face recognition which are overall "about as good […] as developmental prosopagnosics are bad". Police and law enforcement may substantially benefit from SRs in various tasks. These could include secure authentication or surveillance. Most importantly, SRs could significantly improve criminal investigations. For example, SRs have been deployed to investigate the recent New Year's riots in Berlin.* Despite still being only sparsely researched, SRs continue to attract international media attention†,‡. While the UK Metropolitan Police Service was the first to have a designated SR unit§, the first historic SR report was likely that of Eugène F. Vidocq—the famous criminal turned "father of criminology and of the French Police Department," whose methods were adopted by the British Police¶.

While intuitive, the mere suggestion that SRs could improve policing and law enforcement to enhance public safety is of course insufficient. Rather, professionals' abilities should be formally tested to determine their ability for a given task (2–6). A number of studies reporting simulations and empirical data acquired under highly controlled settings (5, 7, 8) suggest that SRs may improve polices' operational proficiency. However, none of the existing studies have to our knowledge examined SRs' performance on processing authentic police, let alone forensic material. Without direct empirical

*https://www.rbb24.de/panorama/beitrag/2023/03/silvesterkrawalle-berlin-erste-anklagen-eingegangen-gericht.html.

†https://www.washingtonpost.com/science/super-recognizer-facial-memory/2021/10/29/4cf80caa-2159-11ec-b3d6-8cdebe60d3e2_story.html.

‡https://www.nzz.ch/schweiz/gesichter-merken-kriminalistik-ist-neuem-phaenomen-auf-der-spur-ld.1703383?reduced=true.

§https://en.wikipedia.org/wiki/Eug%C3%A8ne_Fran%C3%A7ois_Vidocq.

¶https://www.newyorker.com/magazine/2016/08/22/londons-super-recognizer-police-force.

Author contributions: M.R. designed research; M.R. performed research; M.M. contributed new reagents/analytic tools; M.M. and M.R. analyzed data; M.R. contributed research funding; and M.M. and M.R. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: meike.ramon@unil.ch.

Published May 9, 2023.

evidence, the assumption that operations can be improved via SRs remains an indirectly, anecdotally supported hope. Thus, at present, it is still unclear whether there is an actual advantage of SRs for criminal investigations and other police tasks or whether SRs identified via lab-based solutions excel in psychological tests, albeit with no benefit for application in law enforcement.

In 2016, the Cantonal Criminal Police of Fribourg (Switzerland) contacted one of the authors in two ongoing cases of burglary, in which the crimes had been filmed by closed-circuit television (CCTV) installed in the respective bank and jewelry store. To expedite the investigation and perpetrator identification process, the Cantonal Criminal Police sought the help of SRs[#]. This practical request represented a unique opportunity to empirically determine the predictive value of SRs identified with scientifically validated and standardized lab tests for deployment in real-life forensic settings.

However, this opportunity also entailed a number of challenges. First, at the time of the request, when the investigations were ongoing, the ground truth (i.e., perpetrator identity) was unknown. Additionally, inasmuch as individual SRs exhibit varied performance profiles across lab tests (5), they could potentially identify different individuals among suspect line-ups. Therefore, additional CCTV sequences and corresponding line-ups from previously solved cases were solicited from the Cantonal Criminal Police, in order to validate SRs' predictions by taking into account their performance on authentic forensic material with known ground truth. The authentic police material provided was released for scientific purposes, to formally and empirically address the question of whether SRs can improve perpetrator identification as tested in real-life criminal investigation settings. In spring 2017, Swiss police officers arrested the perpetrator, who admitted to having committed both crimes. Thus, for all four authentic CCTV sequences and corresponding line-ups, the ground-truth—i.e., the convicted perpetrator—is now known. This enables us to not only test the performance of SRs on authentic forensic material but also to perform a strong validation of face identity processing tests used to identify SRs.

Advantages and Limits of Lab-Based Assessment of Ability. As of today, a plethora of tests have been developed and used to measure face identity processing, with a large proportion developed to identify individuals with impairments (9). Deviating from the seminal SR report (1), most studies have used a single test to identify SRs. For instance, Bobak et al. (7, 8) used the long version of the Cambridge Face Memory Test (CFMT+ 1). Others have reportedly used "any [qualifying test] in the superrecognizer literature" (10), including, e.g., the Glasgow Face Matching Test (11). Such an indiscriminate approach using varied tests for SR identification increases the risk of undesirable heterogeneity among individuals selected as "SRs". Indeed, tests of face identity processing differ greatly in terms of their sensitivity and reliability (12). The Glasgow Face Matching Test in particular is known to suffer from ceiling effects (2, 11), with "normal" performance exhibited even by highly impaired individuals suffering from developmental or acquired prosopagnosia (13, 14).

To overcome these issues, Ramon (2) recommends three challenging tests assessing face identity perception and recognition and provides cut-off values and criteria to identify SRs. All SRs and control participants in this study completed the three tests in question: the Facial Identity Card Sorting Test [FICST, (15, 17)], the Yearbook Test [YBT, (16, 17)], and the long version of the Cambridge Face Memory Test [CFMT+, (1)]. SRs were identified following Ramon's (2) novel, strict diagnostic framework for lab-based SR identification, achieving high performance in at least two of the three tests.

In the FICST (15, 17), participants are presented with 40 images simultaneously, which depict one of two different individuals. The participants' task is to sort the images into identity groups, effectively telling together and apart the identities included. Thus, the FICST assesses face identity processing through face matching/discrimination, with fewer identity groups, and fewer within-group inclusion errors reflect better performance. The YBT (16, 17) consists of eight trials, each of which involves the matching of five identities depicted by their high school yearbook photographs to their corresponding, 25-y older version among an equal number of same-aged distractors. The test score is determined by the number of correct matches.[‖] Even though not originally developed as a face identity processing test, this test combines unfamiliar face matching across substantial age-related changes and can, thus, be considered as a difficult test suitable for diagnosing SRs. The CFMT+ (1) assesses memory for facial identity via a 3-alternative forced-choice face recognition task. Here, after participants learn target identities, they have to indicate in increasingly difficult trials which of three presented identities corresponds to a previously learned one. Thereby, the test score indicates the number of correct responses across the total of 102 trials. This test was specifically developed to also assess high-performing individuals.

Individuals identified as SRs through the CFMT+ (1) show better performance on experimental procedures resembling face matching of CCTV footage and line-ups (7) or border control scenarios (8). Recent work also demonstrates that SRs identified via tests of face perception and recognition as proposed by Ramon (2) show automatic fixation biases toward faces, which could lay the basis for their superiority in perceptual (18) and delayed matching (3), as well as recognition memory of facial identity (19) as tested with lab-developed, controlled experiments.

A general limitation of existing studies is the typically limited number of SRs reported, which inherently limits their informative value. These fundamental and theoretical aspects, however, are significantly outweighed by two applied concerns. First, to our knowledge, there is no evidence demonstrating that individuals identified as SRs actually show better performance for authentic forensic material. This is a major shortcoming if individuals with strong face identity processing abilities could be highly beneficial in Police tasks and law enforcement. Conversely, authentic forensic material has to date never been used to validate face identity processing tests, thus, limiting the generalizability of their test scores.

A Real-Life Demand Creates a Unique Empirical Opportunity. This study originated from the needs of the Cantonal Police, who were tasked with challenging CCTV-footage–based perpetrator identification for criminal investigation. Responding to their need provided the unique opportunity to address the

[#]Note that originally, only a small sample of (initially 6) SRs had been identified and therefore available to assist. For security reasons, we were not provided any feedback on whether or how their responses were considered.

[‖]The original authors, however, only provided the correct answers for 35 of 40 matches; thus, the maximum score in this test is 35 rather than 40 (17).

aforementioned issues and research gap around SRs and their alleged value in policing. Here, we assess the performance of 73 SRs and 45 control participants identified with the same test material using the same cut-off values (2) on the authentic forensic material provided by the Cantonal Criminal Police of Fribourg. To anticipate our findings, from a test-theoretical perspective, we were also able to validate the test procedures for lab-based SR identification and especially conduct a strong criterion validation using the CCTV sequences provided by the Cantonal Criminal Police of Fribourg. From an applied perspective, we provide the first empirical demonstration of an SR advantage for face identity processing for criminal investigation. Finally, an exploratory analysis investigated interindividual differences between SRs' performance on CCTV footage perpetrator identification. We discuss the results and their implications and provide recommendations for both research and practice.

Methods

Participants. We collected data from 118 participants. 66 were SRs already reported in Ramon (2021) (2) using the recommended test cut-offs and criterion. An additional 52 participants were sampled to allow for a comparison of performance between SRs and individuals with average face processing abilities. However, 7 of the 52 individuals solicited as control participants also met the criteria for SRs and were therefore assigned to the SR group in the following analyses. Participants' mean age was 36.44 ($SD = 9.89$), and 50% of the participants were female. Of all participants, 25.42% were police officers, whereas 74.58% had a different occupation not involving law enforcement. Before viewing the authentic material, the latter group of participants signed a confidentiality agreement provided by the Fribourg Cantonal Police (no such agreement was solicited for law enforcement professionals, who are professionally bound to secrecy).

Material and Procedure. All materials were provided by the Cantonal Police and involved four CCTV sequences and three line-ups. Sequences 1 and 2 represented previously solved cases; sequences 3 and 4 depicted the jewelry and bank robberies under investigation when MR was contacted by the police. Screenshots from the CCTV footage depicting the perpetrator in sequences 3 and 4 and coverage in the media can be found here** and here††, respectively.

The CCTV sequences varied in duration, quality, and camera perspective. Each video could depict a sole perpetrator (two separate bank robberies in sequences 1 and 4; one jewelry robbery in sequence 3), or two perpetrators working together (one case of collaborative pick-pocketing at an ATM in sequence 2). Details concerning the suspect line-up creation were not disclosed. All line-ups displayed 9 suspects of the same (apparent) gender and plausible age variation. Line-up 2 was used for the identification of both perpetrators shown in sequence 2; line-up 3 was used for sequences 3 and 4. None of the sequences contained depiction of physical harm; in all robbery cases, the perpetrator held a handgun.

All participants included in the analysis first completed all three face identity processing tests used as lab-based diagnostic criteria for SR identification (2). For the YBT and CFMT+, test scores reflect the number of correct responses with YBT having a maximum possible test score of 35 and CFMT+ having a maximum possible test score of 102. Contrariwise, performance in the FICST is optimal if individuals sort the presented facial images into two categories without any inclusion errors. Individual scores are computed as (number of groups + number of inclusion errors − 2), where lower values indicate better performance; a score of 0 indicates perfect grouping.

Afterward, they were presented with the four CCTV sequences in the same order. As the second sequence showed two perpetrators, each participant identified five perpetrators in total. The CCTV sequences and line-ups were not modified in any manner; participants were free to explore them as if they were working at/with the police (e.g., they could zoom into a picture, pause, rewind, and rewatch a sequence). After watching a CCTV sequence, participants were presented with the corresponding line-up of 9 suspects next to the CCTV sequence. They were instructed to choose the one representing the closest match to the perpetrator depicted in the CCTV sequence (i.e., they performed a 1-to-9 matching task). Before providing their judgment, participants could rewatch the video as they wished.‡‡

Results

Test Validation. First, to guide not only researchers identifying SRs for scientific purposes but also practitioners who seek SRs for police tasks, we validate the lab tests used to identify SRs in this study and others (3, 18–20). To this end, we examine convergent validity and criterion validity to assess whether all tests measure similar constructs and whether these measured constructs predict performance on external tasks, in this case perpetrator identification in CCTV footage.

If all three tests used in this study, i.e., the FICST, YBT, and CFMT+, measure a similar construct, namely face identity processing abilities, test scores should be at least moderately correlated such that individuals performing well in one test also perform well in the other tests. All correlations reported in Table 1 were computed using Spearman's rank correlation coefficient (ρ) as test scores are not normally distributed and scores for perpetrator identification in CCTV sequences only vary between 0, no correct identifications, and 1, correct identification of all perpetrators. The results are displayed in Table 1 and show medium to high correlations among test scores for FICST, YBT, and CFMT+, with effect sizes in line with previous reports (17). Moreover, Table 1 also displays the rank correlations between test scores and overall performance in perpetrator identification on the four CCTV sequences described above, thereby allowing insights into criterion validity. All three tests show a moderate correlation with the performance on CCTV material such that individuals with higher test scores show better performance on CCTV footage. Taken together, all three tests, i.e., FICST, YBT, and CFMT+, measure the same construct and are related to real-world criteria. This indicates that examining individuals' face identity processing ability with these tests previously proposed as diagnostic measures for SR identification (2) is successful and

Table 1. Correlations among test scores and perpetrator identification on CCTV footage

	1	2	3
1. FICST	–		
2. YBT	−0.40***	–	
3. CFMT+	−0.46***	0.64***	-
4. Perpetrator identification	−0.38***	0.36***	0.43***

*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. Correlations were computed using Spearman's rank correlation.

**https://www.blick.ch/schweiz/westschweiz/bewaffneter-ueberfall-in-freiburg-wer-kennt-diesen-bankraeuber-id5570810.html.

††https://www.24heures.ch/un-voleur-s-empare-de-bijoux-avant-de-fuir-613826312596.

‡‡Note that we did not measure the processing time participants needed to reach their decisions. Within the 1h time slots allocated to participants (initially only SRs tested due to the police request) for forensic material viewing, however, we observed a large degree of variability in processing time—both across observers, as well as trials. Based on our observations, we believe that this variability is likely to reflect individual differences in test-taking behavior and trial difficulty, independently of their ability and motivation.

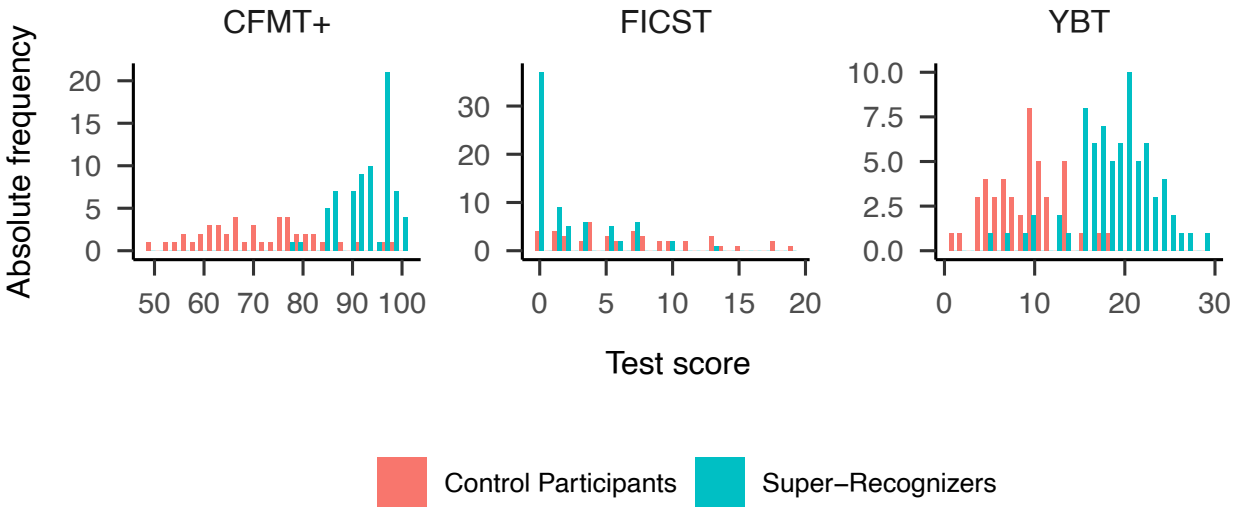

Fig. 1. Distribution of face identity processing test scores.

can predict the performance on authentic forensic material at the group level.[§§]

However, the distributions of test scores depicted in Fig. 1 indicate that compared to the CFMT+ and YBT, the FICST may be less sensitive for detecting high face identity processing abilities. While most SRs obtained low FICST scores (i.e., high performance), so did a large number of control participants. The YBT and CFMT+ on the other hand show more differentiation for high test scores.

Comparing Performance of Super-Recognizers and Control Participants. We next examined whether SRs perform better than control participants in perpetrator identification on CCTV footage. The answers provided by both SRs and control participants are displayed in Fig. 2 where the correct answer is marked by a black rectangle. Overall, sequences show a wide range of difficulty, with average performance ranging from less than 30% correct identification (sequence 3) to almost 90% correct identification for Perpetrator B in sequence 2. However, as participants identified both perpetrators in sequence 2 on the same line-up, this result is likely to be higher than otherwise expected. SRs performed better than control participants for all five instances of perpetrator matching. The largest benefit relative to control participants—a difference of 35%—was observed for sequence 1.

To statistically test the difference in performance between SRs and control participants, we used a generalized linear mixed model assuming a binomial distribution with a logit link function. Perpetrator identification on CCTV sequences (coded as 0—no correct identification, 1—correct identification) served as dependent variable in this model, whereas participants' face identification abilities (coded as 0—control participant, 1—SR) was our independent variable. Moreover, we included random intercepts for participants and CCTV sequences since all participants watched all sequences and provided a judgment for the perpetrator in the video. As expected, SRs perform significantly better at identifying perpetrators on CCTV footage than control participants ($\beta = 0.958$, $CI = [0.542, 2.160]$, $z = 4.394$, $P < 0.001$). These results illustrate that SRs' exceptional performance on tests of face identity processing is accompanied by more accurate perpetrator identification in authentic CCTV footage. This indicates that SRs can be beneficial in perpetrator identification relative to individuals with comparatively lower, i.e., normal, face processing abilities.

Beyond correct identity matches, Fig. 2 also displays which distractors both SRs and control participants selected. Overall, the similarity in SRs' and control participants' response patterns reflect the fact that, in the case of incorrect matches, both groups select largely the same distractors. However, control participants were more strongly attracted by distractors than SRs, most notably for sequence 3. Here, control participants selected one distractor even more frequently than the correct identity, whereas SRs chose both identities equally frequently. Most strikingly, for sequence 4, SRs reliably selected the correct identity, while control participants strongly tended toward choosing a specific distractor.

Perpetrator Identification with Super-Recognizers. Since, at the time of their request, the Cantonal Police of Fribourg sought for help from SRs to solve two ongoing criminal cases, we also assess whether their responses would have actually helped solve them. Given that sequences 1 and 2 were provided as test material to assess SRs' abilities independently of lab-based tests, we examine whether SRs, who correctly matched line-up images to CCTV footage in sequences 1 and 2, also correctly matched line-up images in sequences 3 and 4.

The left of Table 2 shows the relative frequency of correct identifications in sequences 3 and 4 for SRs who either correctly matched all three perpetrators in sequences 1 and 2 or who did not. Performance is better for SRs who show perfect matching in sequences 1 and 2, especially for sequence 3 with an advantage in correct identification of 23.73%.

Since the Cantonal Police suspected that the same perpetrator was responsible for both cases, we additionally examine whether this information would have helped to correctly identify the perpetrator. These results are displayed on the right of Table 2. Overall, SRs show high correct identification rates if the same perpetrator was identified in sequences 3 and 4 with even better identification for SRs who already correctly matched all perpetrators in sequences 1 and 2.

[§§] To check the robustness of our results, we performed the same analyses only for control participant. This served to address the disproportionate SR-to-control ratio and to ensure that our results are not induced by large variances in test scores. We found similar results for convergent and criterion validity with medium to high correlations among tests and medium correlations of tests with performance on CCTV-based perpetrator identification. This indicates that our results are robust even when analyses are restricted to control participants' performance.

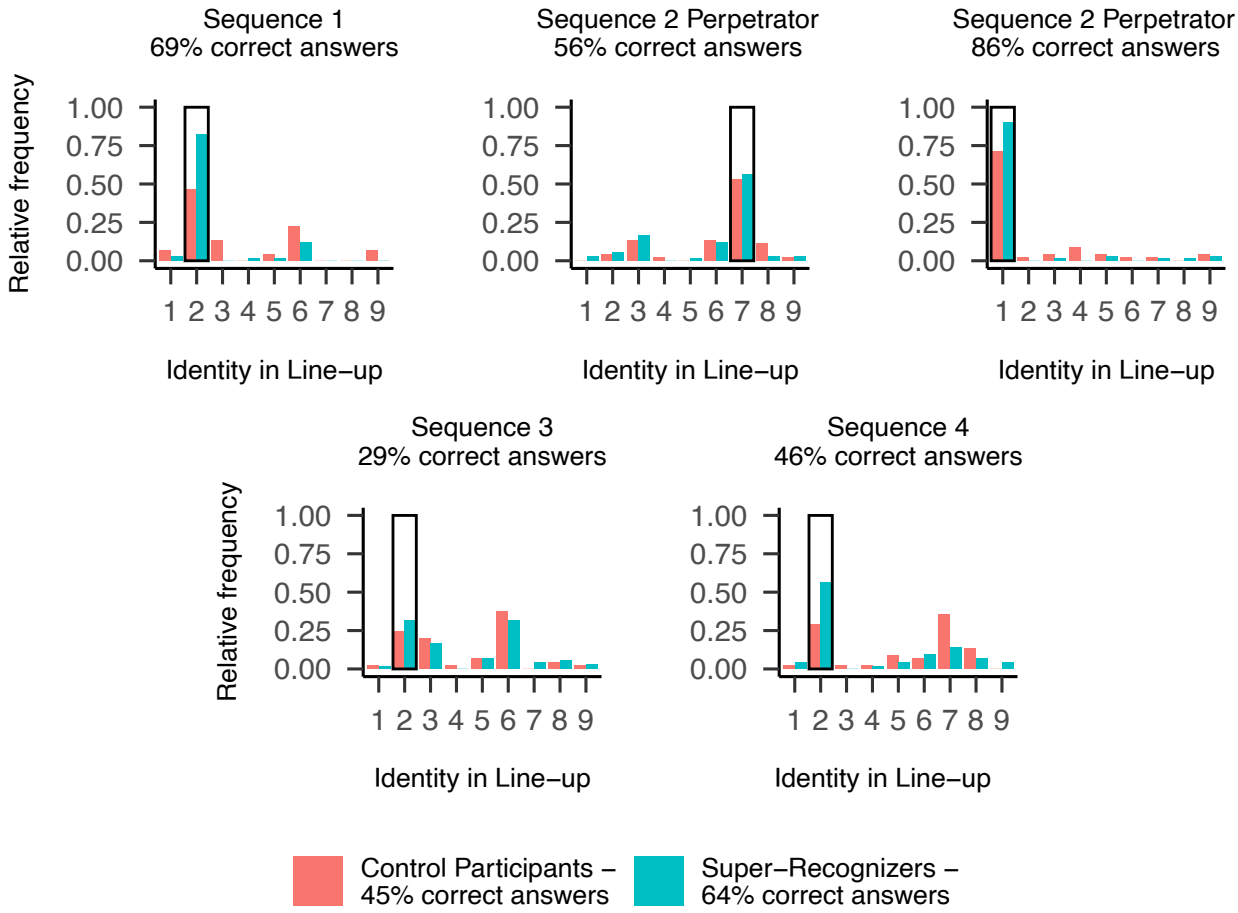

Fig. 2. Perpetrator identification from CCTV footage in corresponding line-ups for both SRs and control participants. *Note*. The correct facial image in the line-up is indicated by a black rectangle.

Thus, following the majority of SRs who already showed perfect performance for sequences 1 and 2 would have been an effective strategy for identifying a likely suspect even in the very difficult scenario of sequence 3. Moreover, additionally considering the information that both CCTV sequences depict the same perpetrator could substantially improve the identification rate, thereby providing even stronger evidence.

Differentiating between Super-Recognizers. When identifying SRs for policing or other tasks, typically, it would neither be feasible, nor of interest to select and permanently deploy very large groups of SRs, but rather to identify a select group of the best-performing individuals. Since typically only lab tests and no authentic material are available for this purpose (with the exception of a bespoke procedure created within the Berlin Police, 2, 4, 21), we explored whether test scores for FICST, YBT, and CFMT+ still predict performance on CCTV footage among SRs or whether individuals who were already identified as SRs all show similar performance. To this end, we compare the performance of SRs on the CCTV sequences who reached SR criteria either in all three tests or only in two of three tests which is according to Ramon (2021) still acceptable to identify an individual as a SR.

The results of this comparison are displayed in the left of Table 3 revealing that SRs reaching respective criteria in all three tests make more correct perpetrator identifications, as compared to SRs who "only" reached corresponding performance levels in two of three tests. This indicates that SRs who exhibited consistently high performance in all three tests perform better on later identifications on CCTV footage.

In a second step, we also took the performance of SRs in sequences 1 and 2 into account since these CCTV sequences were provided as additional testing material for assessing their ability for processing facial identities. As before, we differentiate

Table 2. SR's ability to solve the criminal case based on their performance in already solved cases

		Performance in				Performance in	
Sequences 1 and 2 as test criteria	*N*	Sequence 3, %	Sequence 4, %	Identity decision in Line-ups 3 & 4	*N*	Sequence 3, %	Sequence 4, %
All identifications correct	31	45.16	64.52	Different perpetrators	18	22.22	55.56
				Same perpetrator	13	76.92	76.92
Not all identifications correct	42	21.43	50	Different perpetrators	30	10	50
				Same perpetrator	12	50	50

Table 3. Relative frequency of correct identification by SRs in the CCTV footage

SR test criteria	*N*	Performance in sequences 1 to 4, %	Sequences 1 and 2 as test criteria	*N*	Performance in sequences 3 and 4, %
Criteria fulfilled in three tests	33	71.52	All identifications correct	18	63.89
			Not all identifications correct	15	40
Criteria fulfilled in two tests	40	58	All identifications correct	13	42.31
			Not all identifications correct	27	33.33
Control participants	45	45.33	All identifications correct	8	31.25
			Not all identifications correct	37	25.68

Note: To improve comparability of SR performance, the performance of control participants was added.

between SRs who showed perfect identification in sequences 1 and 2 and SRs who showed less than perfect performance. The right half of Table 3 shows that SRs, who reach respective criteria in all three face identity processing tests and also show perfect performance on CCTV sequences 1 and 2, have by far the best identification rates for sequences 3 and 4 with 63.89% correct identifications. Two groups of SRs show similar identification rates in sequences 3 and 4: SRs who achieve above average lab-based test performance for all three tests proposed as criteria by Ramon (2), but do not perform perfectly in sequences 1 and 2, and SRs who perform worse on tests but perfect on CCTV footage (40% and 42.31%, respectively). These results indicate that SRs' predictive value increases the more consistent their performance is across various tasks.

Discussion

Since their first description by Russell et al. (1), empirical research has focused largely on ways to identify SRs and assess their performance across various experimental settings. Reports of SRs deployed in law enforcement are to our knowledge offered solely via media coverage, where SRs are portrayed as infallible and beneficial for various operative roles. What is missing is empirical data on the real-life utility of lab-identified SRs based on which evidence-informed recommendations can be derived. We bridge this gap between research and practice via three important steps. First, we validate tests proposed as diagnostic criteria for SR-identification (2) on authentic forensic CCTV footage. Second, we assess the performance of 73 lab-identified SRs and 45 control participants on authentic criminal investigation material used for perpetrator identification. Finally, we demonstrate how SRs can help to identify suspects in actual criminal cases.

Concerning the validation of FICST, YBT, and CFMT+, we found that all three tests measure a similar construct, i.e., face identity processing, and that their respective test scores predict identification performance on CCTV footage. Thus, these tests show good construct validity and criterion validity and are suitable to assess individuals' face identity processing abilities, especially when these individuals are selected for similar tasks. Our results align with earlier studies that reported that SRs identified via only a single test excelled at identifying individuals in experimentally created mock CCTV footage (7, 8). Moreover, assessing perpetrator identification ability, we observe that SRs can more accurately identify suspects on authentic police CCTV footage and show even better performance if they already demonstrated high performance on other forensic material.

Importantly, there are still relevant performance differences between individual SRs. Specifically, SRs who met all three (as compared to only two) lab criteria showed higher correct identification rates on authentic CCTV footage. This advantage even increased when their performance on independent CCTV footage was taken into account. Put simply, there are notable differences even among the top of the crop; the more criteria used and met, the better the to-be-expected performance outcome.

Note that here, for our lab tests, as well as the authentic police material used, target identities were always present. Indeed, criminal investigation line-ups may not always contain the target identity in question. Therefore, recently, Boudry et al. (22) systematically investigated the effect of target prevalence on observers' face recognition performance using the same 3-alternative forced-choice memory paradigm as the CFMT (1). Mirroring findings from within the field of visual search (23, 24), the authors report that decreased target occurrence was associated with lower recognition performance. The extent to which SRs are differentially impacted by (changes in) target prevalence requires future investigation.

Recommendations for Researchers and Practitioners. When identifying SRs for scientific (or applied) purposes, we strongly recommend following the recommendations by Ramon (2). For applied purposes, her proposed lab-based criteria should ideally be accompanied by measures designed to mirror the specific tasks, selected individuals are intended to perform. We show that SRs who meet the proposed criteria perform significantly better on authentic forensic material than individuals who do not. Applying these criteria (2) widely would also benefit researchers and practitioners to establish a common definition of the term Super-Recognizer, and prevent misconceptions while realistically managing expectations.

Moreover, our results indicate that individuals scoring above the recommended criteria for all three tests of face identity processing, also show better performance on authentic material as compared to SRs who excelled at "only" two. Thus, if more than needed SRs are identified via lab-based procedures, those excelling across the board should be preferred. Nonetheless, SRs who excel at two of three lab tests still show high(er than normal) performance.

Furthermore, our results indicate that SRs' predictions (e.g., suspect leads) can be improved if their performance on similar tasks is known and considered. We therefore recommend that, in addition to lab-based tests of face identity processing, SRs' abilities should be examined on material resembling their professional task (2, 3, 18, 19). Best on-the-job performance can be expected if SRs are identified using a combined approach that identifies individuals who 1) reach all three of the recently proposed lab-based criteria (2) and 2) also show high performance in tasks that are professionally relevant and that involve authentic material.

Limitations and Future Research Directions. Despite the interesting and important finding for research about and identification of SRs, from an experimental perspective, this work still has obvious limitations. First, only four CCTV sequences involving five perpetrators to identify were available for validating lab-based tests of face identity processing as well as SRs' performance on authentic forensic material. This limits the generalizability of our results both due to only few sequences and due to only one type of authentic material presented to participants: matching a perpetrator via CCTV footage to suspect options presented in a line-up with no time constraints. Moreover, internal consistency may be reduced resulting in lower associations between face identity processing tests and performance in CCTV sequences than typically observed in real-world applications with more trials of authentic material. Nonetheless, we already find strong associations between the tests and performance in perpetrator identification which may be even larger with more trials.

Although experimentally desirable, larger number of sequences and line-ups are typically not feasible, as authentic forensic material cannot be readily made available for scientific research. A greater number of trials allowing investigation of relationships between performance accuracy and response times would also be necessary to determine potential motivational effects, which would be reflected in individuals' response times. Although we believe but cannot ascertain post hoc that response time differences would more likely reflect individual differences in test taking behavior and trial difficulty, two important considerations support this notion. First, the informative value of response times (RTs) varies as function of task type. They are particularly informative in choice tasks with a small number of options and thus high chance level (e.g., 2-alternative forced-choice matching tasks, for a discussion see 13). Additionally, since there are large intraindividual differences in RTs, they are particularly relevant when performance measures are aggregated across a *large number* of trials, across which potential speed-accuracy trade-offs can emerge. In the present study—with only 4 highly variable CCTV sequences in terms of duration and visual information available—we would be reluctant to consider RTs.

Examining the SRs' performance and validating face identity processing tests using authentic material is important for both researchers and practitioners as it connects research on SRs with the application of their abilities in practical tasks. For this reason, existing bespoke applied and scientifically validated solutions such as beSure® (4) should ideally be used by across agencies to allow independent comparisons across means of SR status assessment and SR deployment evaluation (21).

Lastly, the CCTV footage depicted only male Caucasian perpetrators. As both SRs and control participants are likely to be susceptible to demographic effects, a systematic investigation of the effects of gender, age, and race on SRs performance is needed. Future work examining, e.g., the other-race bias (25, 26) among *consistently* lab-identified SRs (in contrast to, e.g., refs. 27 and 28 is needed to determine whether SRs' exceptional face identity processing abilities renders them immune to perceptual biases cf. ref. 29).

Conclusion

The term Super-Recognizer (SR) describes individuals with superior face identity processing abilities. Our study shows that SRs identified via a recently proposed diagnostic framework (2) excel at identifying perpetrators in authentic forensic material. Moreover, SRs are not a monolith: Among their ranks, the more consistent their performance across diagnostic criteria, the better their performance for real-life perpetrator identification was. Thus, deploying SRs in policing and law enforcement tasks requiring face identity processing can be highly beneficial, as evidenced by increased accuracy for forensic perpetrator identification in CCTV footage.

Data, Materials, and Software Availability. Anonymized research data reported (i.e. responses provided by the participants) subject to analysis and the analysis code (30). Authentic police material (videos, images) cannot be displayed or shared. Previously published data (2) were used for this work.

ACKNOWLEDGMENTS. Special thanks are directed to the Cantonal Police of Fribourg (CPF) without which this work would not have been possible: Detectives MB and JB for reaching out and requesting help from SRs in the first place and the CPF for permission to use the provided authentic criminal case material for research purposes. We thank all participants for their support and René Rolvien for the support during collection of control participants' data. M.R. is supported by a Swiss National Science Foundation PRIMA (Promoting Women in Academia) grant (PR00P1_179872).

Author affiliations: [a]Leibniz-Institut für Wissensmedien (Knowledge Media Research Center), 72076 Tübingen, Germany; and [b]Applied Face Cognition Lab, University of Lausanne, 1015 Lausanne, Switzerland

1. R. Russell, B. Duchaine, K. Nakayama, Super-recognizers: People with extraordinary face recognition ability. *Psychon. Bull. Rev.* **16**, 252–257 (2009).
2. M. Ramon, Super-recognizers - a novel diagnostic framework, 70 cases, and guidelines for future work. *Neuropsychologia* **158**, 107809 (2021).
3. J. D. Nador, M. Vomland, M. M. Thielgen, M. Ramon, Face recognition in police officers: Who fits the bill? *Forensic Sci. Int.: Rep.* **5**, 100267 (2022).
4. M. Ramon, S. Rjosk, *beSure® – Berlin Test for Super-Recognizer Identification Part I: Development* (Verlag für Polizeiwissenschaft, 2022).
5. M. Ramon, A. K. Bobak, D. White, Super-recognizers: From the lab to the world and back again. *British J. Psychol.* **110**, 461–479 (2019).
6. R. Moreton, G. Pike, C. Havard, A task- and role-based perspective on super-recognizers: Commentary on 'super-recognizers: From the lab to the world and back again'. *British J. Psychol.* **110**, 486–488 (2019).
7. A. K. Bobak, A. J. Dowsett, S. Bate, Solving the border control problem: Evidence of enhanced face matching in individuals with extraordinary face recognition skills. *PLoS One* **11**, e0148148 (2016).
8. A. K. Bobak, P. J. B. Hancock, S. Bate, Super-recognisers in action: Evidence from face-matching and face memory tasks. *Appl. Cogn. Psychol.* **30**, 81–91 (2016).
9. R. Russell, G. Chatterjee, K. Nakayama, Developmental prosopagnosia and super-recognition: No special role for surface reflectance processing. *Neuropsychologia* **50**, 334–340 (2012).
10. P. J. Phillips *et al.*, Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 6171–6176 (2018).
11. A. M. Burton, D. White, A. McNeill, The glasgow face matching test. *Behav. Res. Methods* **42**, 286–291 (2010).
12. A. K. Bobak *et al.*, Data driven approaches can tell us as much about the processes underlying face cognition as the quality of the available tests and data sets allows. Where do we go from here? (2021).
13. M. C. Fysh, M. Ramon, Accurate but inefficient: Standard face identity matching tests fail to identify prosopagnosia. *Neuropsychologia* **165**, 108119 (2022).
14. D. White, D. Rivolta, A. M. Burton, S. Al-Janabi, R. Palermo, Face matching impairment in developmental prosopagnosia. *Q. J. Exp. Psychol.* **70**, 287–297 (2017).
15. R. Jenkins, D. White, X. Van Montfort, A Mike Burton, Variability in photos of the same face. *Cognition* **121**, 313–323 (2011).
16. M. Bruck, P. Cavanagh, S. J. Ceci, Fortysomething: Recognizing faces at one's 25th reunion. *Mem. Cogn.* **19**, 221–228 (1991).
17. L. Stacchi, E. Huguenin-Elie, R. Caldara, M. Ramon, Normative data for two challenging tests of face matching under ecological conditions. *Cogn. Res.: Princ. Implic.* **5**, 8 (2020).
18. J. D. Nador, M. Zoia, M. V. Pachai, M. Ramon, Psychophysical profiles in super-recognizers. *Sci. Rep.* **11**, 13184 (2021).
19. J. D. Nador, T. A. Alsheimer, A. Gay, M. Ramon, Image or identity? Only super-recognizers' (memor)ability is consistently viewpoint-invariant. *Swiss Psychol. Open: Offici. J. Swiss Psychol. Soc.* **1**, 2 (2021).
20. M. Linka, M. D. Broda, T. Alsheimer, B. de Haas, M. Ramon, Characteristic fixation biases in super-recognizers. *J. Vision* **22**, 17 (2022).
21. M. Ramon, S. Rjosk, *European Recommendations for the Protection of Public Spaces against Terrorist Attacks* (2021).
22. L. Boudry, J. D. Nador, M. Ramon, Determinants of face recognition: The role of target prevalence and similarity (2023).

23. J. M. Wolfe , et al., Low target prevalence is a stubborn source of errors in visual search tasks. *J. Exp. Psychol.: General* **136**, 623–638 (2007).
24. J. M. Wolfe, M. J. Van Wert, Varying target prevalence reveals two dissociable decision criteria in visual search. *Curr. Biol.* **20**, 121–124 (2010).
25. A. M. Megreya, D. White, A. M. Burton, The other-race effect does not rely on memory: Evidence from a matching task. *Q. J. Exp. Psychol.* **64**, 1473–1483 (2011).
26. J. W. Tanaka, M. Kiefer, C. M. Bukach, A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study. *Cognition* **93**, B1–B9 (2004).
27. S. Bate , et al., The limits of super recognition: An other-ethnicity effect in individuals with extraordinary face recognition skills. *J. Exp. Psychol. Hum. Percep. Perf.* **45**, 363–377 (2019).
28. D. J. Robertson, J. Black, B. Chamberlain, A. M. Megreya, J. P. Davis, Super-recognisers show an advantage for other race face identification. *Appl. Cogn. Psychol.* **34**, 205–216 (2020).
29. E. Marini, C. Lavitola, C. Casco, L. Battaglini, L. Ronconi, Audiovisual rhythmic stimulation within the beta band improves the perception of crowded stimuli. **50**, 188–189 (2021).
30. M. Mayer, M. Ramon, Supplemental material for "Improving forensic perpetrator identification with Super-Recognizers". OpenScienceFramework (2022). https://osf.io/u7ge2/.