



**HAL**  
open science

# An attempt to identify predictive features among Islamist radicals: Evidence from machine learning

Michaela Pfundmair, Jeremy Lefort-Besnard

## ► To cite this version:

Michaela Pfundmair, Jeremy Lefort-Besnard. An attempt to identify predictive features among Islamist radicals: Evidence from machine learning. 2020. hal-03042809

**HAL Id: hal-03042809**

**<https://hal.science/hal-03042809>**

Preprint submitted on 7 Dec 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**An attempt to identify predictive features among Islamist radicals:  
Evidence from machine learning**

Michaela Pfundmair<sup>1,\*</sup> and Jeremy Lefort-Besnard<sup>2</sup>

<sup>1</sup>Federal University of Administrative Science

<sup>2</sup>National Tsing Hua University

\*Corresponding author: [michaela.pfundmair@hsbund-nd.de](mailto:michaela.pfundmair@hsbund-nd.de)

Word count: 6,511 words

**Funding details:** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. **Disclosure statement:** The authors declare that there are no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. **Biographical note:** Michaela Pfundmair is professor of psychology at the Faculty of Intelligence at the Federal University of Administrative Sciences in Germany. Her main line of research is about radicalization and social exclusion. Jeremy Lefort-Besnard is postdoc at the Behavioral Informatics and Interaction Computation Lab at NTHU in Taiwan. He works on predictive models based on behaviors and brain activities using machine-learning tools. **Data availability statement:** All data and analysis scripts of the present study are openly and readily accessible to the reader online (<https://bit.ly/2Zx9hpV>).

### **Abstract**

Past research has made substantial efforts to understand the radicalization process. However, to date, quantitative research in this area is undeveloped. Therefore, we used a machine learning approach to identify predictive variables in the context of Islamist radicalization contrasting radicalizing individuals with a meaningful control group. Our algorithm i) predicted the radical status with an accuracy as high as 80.36% and ii) uncovered the most predictive features of radical Islamists among the explored variables. A tendency to withdraw, being male, and a high lack of perspective were highly indicative of being radical, while not feeling excluded, the mosque or friends being the source of radicalization, and low authoritarianism were features highly indicative of not being radical. Contributing in a new way to terrorism research, our study might add to the understanding of radicalization and inspire the work of practitioners to improve detection and prevention efforts.

*Keywords:* Islamist; radicalization; predictive; sparse modeling; machine learning

## **An attempt to identify predictive features among Islamist radicals:**

### **Evidence from machine learning**

Terrorism is a topic which does not seem to age. Although it existed way before this incident, the Western world became increasingly aware of it from the 9/11 attacks on which, in the same breath, produced a great mass and also milestones in this area of research. Terrorism can be defined as an act of violence which aims to achieve behavioral change and political objectives by creating fear in larger populations (Doosje et al., 2016). However, from a psychological point of view, terrorism itself is less mysterious than the way to terrorism. Most people who become terrorists in the Western world face a radicalization process which involves the adoption of an extremist worldview that legitimizes the use of violence to reach their goals (Porter & Kebell, 2011). The question on how radicalization develops in detail is not clearly answered to date.

### **Features of Radicalization**

Several models on radicalization have been developed in the past. One model, for example, centers the quest for significance in the radicalization process: It assumes that the radicalization process starts with an individual whose goal to reach significance is activated. Then, the individual identifies terrorism as appropriate means to reach this goal and, finally, the goal becomes dominant over alternative ones which are incompatible with terrorism (Kruglanski et al., 2014). In another model, three chronological phases are distinguished (Doosje et al., 2016): 1) A sensitivity phase which involves factors within the person that drives the radicalization process, e.g., feelings of insignificance or personal uncertainty. 2) A group membership phase in which mutual commitment is central. And 3) an action phase in which radicalizing individuals turn to using violence. Although such models are helpful in structuring the full radical development, they are often too unspecific to identify individual features of radicalization as they remain on a meta-level.

Besides such models, research has investigated single characteristics in the radicalization process: For example, a study which investigated individuals who radicalized in the United States using the PIRUS database found four predictors of violent extremism in multivariate models: stable employment, radical peers, mental illness, and criminal record (LaFree et al., 2018). However, the authors concluded that, although informative, these factors should not be seen as complete profile of a violent extremist because they are based on a restricted dataset. Another study which compared right-wing, left-wing, and Islamist extremists in the United States using another public database, revealed Islamist perpetrators to be younger, less likely to have a college education, and to be rarely married (Chermak & Gruenewald, 2015). Although such individual findings are intriguing as they give evidence for very concrete characteristics, they can only provide snippets of a range of diverse mechanisms which play a role in the radicalization process.

Combining both advantages (completeness and concreteness), a recent approach tested clusters of features in a comprehensive model of Islamist radicalization using a case study method (Pfundmair et al., 2019). The model suggested individual preconditions to precede a gradual increase of individual processes, group processes and cognitive processes resulting in violent attitudes, intentions and, ultimately, actions. Specifically, the individual preconditions included being male with an age in the early twenties, and having a second-generation migration background and biographical cuts like critical life events or social exclusion. The individual processes covered pursuing individual needs like transcendence, significance, self-esteem and control. On the group level, processes like group identification, prejudice, polarization, perceived group threat, collective emotions and informative influence fueled the radicalization process. Finally, cognitive processes which made people increasingly comfortable with the idea to apply violence included desensitization and dehumanization. On personality factors like authoritarianism or social dominance orientation, the study did not find a consistent pattern. The data used in this work appears generally suitable to narrow

down the radicalization process, however, it was mainly analyzed descriptively or by simple mean differences. A more elaborate statistical processing of this data would allow to identify predictive features among Islamist radicals.

### **Problems in the Methodological Approaches to Understand Radicalization**

Although previous work has made substantial efforts to understand the process of radicalization, the current state of research is not yet able to present a consensual explanation of the phenomenon. According to Sageman, “we are no closer to answering the simple question of what leads a person to turn to political violence” (Sageman, 2014, p. 565; but see, for example, Schmid, 2014). One main problem of the research on terrorism and radicalization is its methodological approach. The use of statistics is a particular area of concern: Although there was a slow upward trend in recent years, 78.1% of work on terrorism and radicalization does not use statistics at all (Schuurman, 2020). Moreover, elaborate statistical approaches are rare. Thus, quantitative research in this area remains undeveloped. The current work aimed to address this gap.

### **The Current Work**

In view of potential benefits from allowing the data to “speak for themselves”, we revisited the traditional approach of studying Islamist radicalization. The current research focused on building an algorithmic framework to achieve individual level predictions. Specifically, by using a machine learning approach in a sample of actual Islamist radicals who were contrasted with a meaningful control group, we aimed to identify predictive variables in the context of Islamist radicalization. Importantly, our goal was not to create a model of radicalization or to classify terrorists. Instead, we aimed to use a data-driven approach to identify variables that are meaningful among radical Islamists while focusing on a comprehensive set of variables that revealed to be relevant in previous research. Thus, the present investigation extends previous approaches by moving from a traditional explanatory setting to a predictive approach. This analysis framework allows to focus on the predictability

of each variable at the level of single individuals. To the best of our knowledge, the current study is the first to apply such an elaborated statistical approach to a sample of radicals.

Two aspects should be noted: First, the database for this approach was data collected in Pfundmair et al. (2019, Study 1). Thereby, we benefited from including features in our statistical model that were gained to obtain a complete overview of the radicalization process while being sufficiently concrete. We included all collected variables even if they did not find their way into the final model. Second, the current work focused on Islamists who had positive attitudes towards violent acts but had not (yet) conducted a violent act on behalf of their beliefs. Consistent with research on attitude and behavior (Glasman & Albarracín, 2006), radical attitudes do not necessarily result in radical action (Borum, 2011) why the development of ideologies and beliefs and the development of engaging in terrorism needs to be separated (McCauley & Moskaleiko, 2017). Accordingly, this study investigates predictive features of the former.

## **Methods**

### **Sample**

In the current work, we used data collected by Pfundmair and colleagues (2019, Study 1). The sample consisted of 75 people suspected of Islamist radicalization who were investigated by a German state office of criminal police specialized for deradicalization. They were reported to this office by government agencies (e.g., police units or state prisons) or by concerned civilians via hotline.

Police professionals who had personal contact with all of these subjects for a certain amount of time in which deradicalization efforts had taken place classified whether the subjects were indeed radical or not. In total, 56 subjects were classified as radical. Radicals were classified as such as soon as they had exhibited positive attitudes towards violent acts with Islamist background; importantly, these had not (yet) executed a terrorist criminal act. Although appearing as such from the outside, the remaining 19 subjects showed no sign of

advocating violence in support of Islamism during deeper investigation by the police, and thus, were classified as non-radical. The distribution of age and sex in the sample was homogeneous (see Fig. 1).

[Figure 1 about here]

### **Data Acquisition**

Two ways of data acquisition were applied: 1) Variables (sociodemographic and biographic factors) were extracted from each subject's existing case file. 2) Further variables (biographic and personality factors, individual, group and cognitive processes, outcomes) were assessed by ratings of the involved police professionals. (Note that the repertory of variables originated from the theoretical rational of Pfundmair et al., 2019.) Specifically, each subject was rated by that professional who was in charge of the particular case and, thus, personally knew the subject. All police professionals were educated about the items and underlying constructs by two psychologists beforehand. Then, they assessed the subjects on the different variables through a questionnaire that primarily used three-point rating scales with behavioral anchors to simplify coding.

**Variables.** Given that not all information was available for each subject, in this study, we discarded variables with too many missing entries as it would be deleterious for the statistical analysis (for details, see Supp. Tab. 1). Using this approach, a total of 19 features (for details and descriptive statistics, see Supp. Tab. 2 and 3) were under study in the current work: radical condition (radical; non-radical), age at radicalization, sex (male; female), migration background (no migration background; third generation; second generation; lives for more than two years in Europe; lives for less than two years in Europe), refugee status (refugee; no refugee), convert status (convert; no convert), marital status (single; unmarried with partner; married/state or Islamic law; divorced; widowed), residential situation (with parents; alone; with partner; shared apartment; children's home), critical life-event (critical life-event [e.g., fleeing, parental divorce]; no critical live-event), criminal record (no criminal



record; one-time crime; more than one crime), violent crime (no violent crime; one-time violent crime; more than one violent crime), social exclusion (never experienced any social exclusion, is well integrated into his or her environment; experienced moderate social exclusion or was excluded in a specific area of life [e.g. at school], the exclusion was short-term; experienced permanently extreme social exclusion in different areas of life [outsider at school, no friends, total isolation], the exclusion was long-term), lack of perspective (seems to have a positive future or has a good chance for a successful life [e.g., work, partnership]; seems to have an uncertain future or would have a chance if trying; seems to have a poor future or has no chance for a successful life [e.g., long-term unemployment, threatening imprisonment]), authoritarianism (rejects every form of authority, refuses to subject to authorities [e.g. disobeys his parents or the state]; accepts and obeys authorities [e.g. school, police] or considers them as legitimized and necessary; demands higher level of authority, is entirely willed to submit himself/herself completely, considers every authority as indispensable), visit to Syria or other (no; yes), agitation (no; yes), change of appearance (no; yes), withdrawal (no; yes), and source of radicalization (family; partner; friends; school; mosque; environment; prison; Internet).

### **Statistical Analyses**

**Identifying predictive features: sparse logistic regression.** We applied a modeling technique that emphasizes both prediction performance and automatic identification of the most relevant features to predict a radical status along the explored variables. That is, we wanted to fit an algorithm that relates the included variables to the extent of being or not being a radical Islamist, with the aim of accurately predicting the response for future individuals. To achieve this goal, we capitalized on the pattern-learning algorithm sparse logistic regression (Friedman et al., 2001). This procedure essentially gains knowledge on two aspects: On the one hand, this predictive learning algorithm estimates the separating hyperplane (i.e., a linear function) that distinguishes radical from non-radical participants. The

outcome  $y$  is defined by being a radical Islamist (1) or not (0). On the other hand, the imposed sparsity constraint identifies a minimal subset of features that is most informative about the differences between individuals. While computing a vector of weights associated with the features is similar to linear regression, this approach aims to reduce the weights of features that have little discriminatory value to exactly zero (a feature's weight equal to zero has no effect on the prediction outcome). This procedure results in a subselection of features which have high joint discriminatory power to separate radicals from non-radicals. In other words, the quantitative investigation detected the subset of features that was most predictive of the radical status. In this way, sparse logistic regression extends previous radicalization investigations by automatic variable selection (Hastie et al., 2015). As a result, our modeling, obtained from the sparse logistic regression, is much easier to interpret. Note that we refrained from computing more sophisticated regressions with non-linear interaction terms because these model extensions would require more than twice the sample size to obtain model fits of equal quality.

The sparsity constraint was imposed in form of an  $l_1$  regularization. Such a constraint in the optimization objective automatically detects relevant features “on-the-fly” during model estimation. The  $l_1$  penalty term, calibrated by the hyper-parameter  $\lambda$ , is designed to control the parsimony criterion and its shrinkage regularization on the learned model weights. This penalized (negative) log likelihood of the logistic regression objective is given by:

$$-\frac{1}{N} \sum_{i=1}^N \log \left( 1 + e^{-y_i f(x_i; \beta_0, \beta)} \right) + \lambda \|\beta\|_1$$

where  $x_i$  represents a given individual's information,  $y_i$  is his/her radical status (0 if not radical, 1 if radical),  $\beta_0$  is the intercept, and  $\beta$  is the weight attached to each feature, the right part of the equation corresponds to the  $l_1$  penalty term controlled by the hyper-parameter  $\lambda$ . The feature selection behavior depends on the choice of this tuning parameter. Indeed, the sparse logistic regression shrinks the coefficient estimates more toward zero and performs

always more aggressive variable selection with increasing  $\lambda$ . The hyper-parameter selection was based on the data in a principled fashion using nested cross-validation. In a common grid of candidate parameter choices, the value of  $\lambda$  was varied logarithmically from -3.5 to 1.0 in log-space with 25 steps. The quantitative investigation detected the subset of features that was most predictive of the radical status.

Our sample included more radical ( $n = 56$ ) than non-radical individuals ( $n = 19$ ). Therefore, we handled class-imbalance using up-sampling given that the majority class was inferior to a third of the minority class. We further detailed this analysis with an examination of the learning curve to assess the predictive model performance as a function of increasing sample size. To this end, we computed the prediction accuracy for the sparse logistic regression for predicting the radical status in a series of increasing patient subsets for model training. Finally, we computed the confusion matrix for each classification algorithm to allow visualization of its performance.

**Computing engine.** Python was selected as scientific computing engine. Capitalizing on its open-source ecosystem helps enhance replicability, reusability, and provenance tracking. Scikit-learn (Pedregosa et al., 2011) provided efficient, unit-tested implementations of state-of-the-art statistical learning algorithms (<http://scikit-learn.org>).

### **Open Practice Statement**

The study was not formally preregistered. All data and analysis scripts of the present study are openly and readily accessible to the reader online (<https://bit.ly/2Zx9hpV>). The materials have been made available in the Supplementary of Pfundmair et al. (2019). All measures, manipulations, and exclusions in the study are disclosed.

## **Results**

### **Informative Features for Detecting Radicals**

A sparse logistic regression was used to automatically identify the subset of features that is most informative about telling whether individuals are not radical versus radical (see

Fig. 2 and Supp. Tab. 4). In our sample, 10 features contributed for predicting radicals, while 12 features contributed for predicting non-radicals; all other features did not contribute (0.0).

The 10 features indicative of radical individuals included being male (1.06), showing withdrawal tendencies (1.27), having immigrated more than two years ago (0.52), living alone (0.08), living in a shared apartment (0.05), living with a (married) partner (0.28), having committed a one-time violent crime (0.07), feeling highly socially excluded (0.24), having a high lack of perspective (1.03), and the Internet being the source of radicalization (0.55). Thus, the heaviest weights for indicating a radical individual were a tendency to withdraw, being male, and a high lack of perspective.

The 12 features indicative of non-radical individuals included having experienced a critical life event (-0.38), having already visited Syria or a similar country (-0.44), being a second-generation migrant (-0.11), having a partner without being married (-0.27), living in a refugee home or with parents (-0.17 and -0.37 respectively), not having any criminal record at all (criminal -0.22 or violent crime -0.50), having low feelings of social exclusion (-1.13), being low in authoritarianism (-0.62), and friends or the mosque being the source of radicalization (-0.62 and -0.91 respectively). Thus, the heaviest weights for indicating a non-radical individual were having low feelings of exclusion, the mosque being the source of radicalization, followed equally by friends being the source of radicalization and being low in authoritarianism.

[Figure 2 about here]

Moreover, we used a parsimony-inducing learning algorithm to search through the array of features and extract the most informative subset of features for predicting radicalization. Therefore, we plotted profiles of the classifier coefficients of the features and the decreasing parsimony constraint of this statistical model (here represented as the increasing number of features automatically selected) along an automatically selected model ( $\lambda = 0.75$ ; see Fig. 3). In sum, the features “withdrawal tendencies”, “being male,” and “high

lack of perspective” quickly uncoupled the other features for being highly predictive of radical individuals, while the two features “low feelings of exclusion” and “mosque being the source of radicalization” quickly cut off from the other variables for being highly predictive of non-radical individuals.

[Figure 3 about here]

### **Additional Analyses**

We detailed these analyses with further examinations of the data. First, we computed the learning curve. A learning curve is a representation of performance of the learning algorithm (here a sparse logistic regression) with always bigger subsets of the data at hand. Performance in the training set (the set used to fit and assess the algorithm) higher than performance in the testing set (the set exclusively used to assess the fitted algorithm) is to be expected since the training score can dramatically overestimate the test score. Indeed, the algorithm learns from the training data and therefore is optimistically biased while the test score represents the efficiency of the model applied to unseen data. While recruitment of more participants would likely further improve the performance of the learning algorithms, we can observe that gathering 50 subjects is a good tradeoff between the needed amount of data and achieving a fair classifier performance (see Fig. 4). Thus, our sample size seemed satisfactory.

[Figure 4 about here]

Second, we generated a confusion matrix. The confusion matrix depicts the percentage of correct classification and misclassification. 83.3% of radical participants and 75.0% of non-radical participants were correctly classified by the model (i.e., the sparse logistic regression). On the other hand, 16.7% of radical participants and 25.0% non-radical participants (false alarm) were misclassified. Thus, in each case, the percentage of correctly classified participants was higher than the percentage of misclassified participants, suggesting a good performance of the model (see Supp. Fig. 1).

Finally, the 90% bootstrapped confidence intervals were computed (see Fig. 5) to provide a principled estimate of the statistical quality of each variable importance in the prediction of a radical status. The 90% bootstrapped confidence intervals were calculated by fitting the sparse logistic regression to 1000 bootstrapped samples made from our original sample. Bootstrapped samples can be considered as generated datasets and represent quite good approximations for population parameter (Efron & Tibshirani, 1994). In our study, bootstrapping was used to provide a principled estimate of the statistical quality of the feature importance for predicting a radical status. In sum, each feature highly weighted in our automatically selected model to predict radicals or non-radicals displayed at a one-sided confidence interval, especially the feature “withdrawal tendencies”.

[Figure 5 about here]

### **Discussion**

Addressing the lack of well-controlled empirical studies that rely on statistical techniques (Schuurman, 2020), our data-driven analysis extracted quantitative insights in the context of Islamist radicalization. Adopting a statistical learning approach, we aimed at reconciling previous explanatory research. Specifically, in the current study, we used a reliable machine learning classifier to identify predictive variables of a radical status. In a nutshell, we found two patterns: First, the most predictive features for being a radical Islamist were a tendency to withdraw, being male and a high lack of perspective. These results also suggest conclusions in correlative terms. For example, the more an individual experienced a lack of perspective, the more likely he or she was radical. Second, subjects who did not feel excluded, who were approached with radical ideas in the mosque or by friends and who were low in authoritarianism were particularly resilient to radicalization. In correlative terms, this means, for example, that the less people felt excluded, the higher was the chance that they did not become radical.

### **Predictive Features for Being Radical**

The identified conditions that pushed a radical status fit to previous findings. Withdrawal tendencies as observed in the current study are a typical feature of radicalization as isolation is characteristic for terrorist cells in which its members only trust one another (McCauley & Moskaleiko, 2008; Porter & Kebell, 2011). Under isolation, group members become more dependent on one another which triggers group processes like polarization. This is because, under these circumstances, the information available is selective and inward-looking; thus, it is easy to create an own reality. Some authors have described this as playing out an own “fantasy war” (Ferracuti, 1990). Also, the Internet makes it easier for individuals and groups to withdraw from the world and create an extremely restricted social reality (Koomen & van den Pligt, 2016). Our findings suggest that this withdrawal is an important – and interestingly, clearly visible – aspect which can help predict Islamist radicalization.

With a male gender being another important predictive feature, our findings comply with previous studies that found the vast majority of radicals to be male (Kebell & Porter, 2012). This gender effect might occur as terrorist movements are linked to motives which may be particularly attractive for males. On the one hand, it has been suggested that terrorism may serve the goal to regain masculinity (Ferber & Kimmel, 2008). For example, a radical white supremacist accordingly explained that he felt weak and humiliated without his radical group, whereas inside the group, there were “men, dressed in camouflage and armed with rifles and pistols, [who] shouted, ‘I am strong!’” (Stern, 2003, p. 18). On the other hand, it has been argued that the desire for adventure and glory may attract radicals: Radical groups not only offer bonds of great solidarity and intimacy but also “open up a world of exhilarating action, violence, intrigue and drama” (Cottee & Hayward, 2011, p. 979).

The third highly predictive feature for being radical revealed to be a lack of perspective. Previous research has provided several hints that this feature might play an important role in the radicalization process. For example, in a study analyzing 172 countries between 1970 and 2006, a link between domestic terrorism and minority group economic

discrimination – a combination of employment discrimination, unequal access to educational, social and health services, housing segregation and a lack of economic opportunities available to the rest of society – was found (Piazza, 2011). This alienation from the system seems to induce distrust of state institutions and authority and make people more susceptible to radical ideas. Consistently, case studies acknowledged alienation to be a driving force to join extremist organizations (Wiktorowicz, 2005). Previous models have subsumed this point under the term grievance (Hafez & Mullins, 2015) or sensitivity (Doosje et al., 2016). However, to the best of our knowledge, this study is the first to use a purely quantitative approach to identify a lack of perspective as one of the most predictive conditions of a radical Islamist status.

### **Predictive Features for Being Non-Radical**

The current study also identified advantageous conditions for being non-radical. These fit to insights of previous work. The most predictive feature for a non-radical status were low feelings of social exclusion. The state of social exclusion is accompanied by tremendous psychological stress, feeling deprived in basic needs for belonging, self-esteem, control, and meaningful existence (Williams, 2007). Previous work has suggested that it plays an important role in radicalization (e.g., Pfundmair, 2019; Weight-Neville & Halafoff, 2010). Reversing this reasoning, it seems plausible that low levels of exclusion can protect potentially susceptible individuals from radicalization as suggested in the current work. Consistently, in the past, social inclusion has shown to work preventively and even deradicalizing: For example, demobilization strategies for the Palestine Liberation Organization included incentives for getting married and forming a family (Dechesne, 2011). Similar efforts were made to break up other radical organizations like the Red Brigades in Italy or the FARC in Columbia which turned out quite effective (Horgan & Braddock, 2010). Our study provides empirical evidence that avoidance of social exclusion or even social inclusion of at-risk individuals might be effective to prevent Islamist radicalization.



In 2004, Sageman found that groupings of friends around mosques provided key opportunities for recruitment (Sageman, 2004); several years later, he claimed that the Internet had dramatically transformed the structure and dynamic of terrorism, and that communication and inspiration shifted from face-to-face interactions to interactions on the Internet (Sageman, 2011). Sageman even suggested the Internet to serve as “virtual glue” to an otherwise “leaderless Jihad”. Indeed, the Internet facilitates sharing of radical ideas, enhances social networks and even helps to inspire violent action (Southers, 2013). This offline-to-online development might also be visible in our data in which radical Islamist ideas transported through friends and the mosque did not contribute to the radicalization process. The current findings even suggest that face-to-face interactions might not only be unsuccessful to radicalize individuals in these times but even counterproductive.

Finally, low authoritarianism as protective factor against Islamist radicalization also seems to be consistent with previous research. Authoritarian individuals follow traditional standards and values, respect authority and direct their aggression against targets sanctioned by the latter. This personality trait probably originates from feelings of uncertainty (Koomen & van der Pligt, 2016). Authoritarianism is positively related to a number of factors that facilitate a radical mindset, for example, outgroup antipathy (Altemeyer, 1998), prejudice (Krauss, 2002), and cultural conservatism (Van Hiel et al., 2004). It has also been explicitly shown to moderate radicalization (Chirumbolo, 2002; Koehler, 2014). For example, people in Lebanon who were high in authoritarianism supported acts of terrorism against the United States more strongly than those who were low in authoritarianism (Henry et al., 2005). As a result, protection against a radical mindset through a personality equipped with low authoritarianism seems plausible.

### **Limitations and Future Research**

Combining a machine learning approach with a sample of actual radicals and a meaningful control group, our study paves the way to integrating and amending quantitative

insights to the existing research on radicalization. However, some limitations have to be considered.

First, the current study focused on Islamist radicals. Although all radical groups share basic characteristics, from experiencing problems in society to embracing an ideology that legitimizes violence (see Doosje et al., 2016), a generalization to radicalization processes in other ideologies seems not warranted at this point. Also, whether or not the current subjects' radical opinions will turn into radical actions (McCauley & Moskalenko, 2017) or, in other words, whether these people will actually become terrorists or not, cannot be answered at this point.

Second, the mode of data collection is open to criticism. The data was coded by police professionals and thus third-hand raters, mainly on the basis of behaviorally anchored rating scales that served as proxies for the respective underlying constructs. Thus, we cannot exclude the possibility that certain aspects were overlooked or simply not visible for the police professionals.

Third, although we tested a relatively high number of features in our radicals, this list was certainly not exhaustive. Using data of Pfundmair et al. (2019), we were only able to include those pieces of information that were previously collected. Moreover, these were narrowed down to variables with enough data points. Thus, it is more than likely that there are more factors that may predict the emergence and non-emergence of a radical status. Nonetheless, the here identified variables are still valid and may be helpful – both to detect possibly critical features and developments among people who are at risk to radicalize, and to guide future prevention actions.

Finally, it is important to note that our results are only preliminary and need to be confirmed by future research. Nevertheless, to ensure that learning was carried out in the current work, several complementary analyses were performed: i) The learning curve showed that, starting from 50 subjects, a fair classifier performance was obtained suggesting that the

model extracted heuristic value from the sample. ii) The estimated bootstrapped confidence intervals were computed, and all variables automatically highlighted as very predictive by the model that obtained one-sided confidence intervals suggesting consistency in the results. iii) The confusion matrix exhibited a higher percentage of correctly classified participants. Moreover, to get a valid measure of predictive accuracy, we used a combination of i) a simple model (limited to additive effects between the studied input variables and thus less data-hungry), ii) sparsity (limiting the risk of overfitting), and iii) nested cross validation (ensuring that the pattern-learning process was independent of any knowledge about the data).

### **Conclusion**

Contributing in a new and, according to recent opinions (e.g., Schuurman, 2020), desired way to radicalization research, our study might add to the understanding of Islamist radicalization by providing predictive features of a radical status. It may contribute to support the work of practitioners to identify subjects more prone to sink into Islamist radicalization, and develop means of prevention.

### References

- Altemeyer, R. A. (1998). The other “authoritarian personality.” In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 30, pp. 47-91). Academic Press.  
[https://doi.org/10.1016/S0065-2601\(08\)60382-2](https://doi.org/10.1016/S0065-2601(08)60382-2)
- Borum, R. (2011). Radicalization into violent extremism I: A review of social science theories. *Journal of Strategic Security*, 4(4), 7-36. <https://doi.org/10.5038/1944-0472.4.4.1>
- Chermak, S., & Gruenewald, J. A. (2015). Laying a foundation for the criminological examination of right-wing, left-wing, and Al Qaeda-inspired extremism in the United States. *Terrorism and Political Violence*, 27(1), 133-159. <https://doi.org/10.1080/09546553.2014.975646>
- Chirumbolo, A. (2002). The relationship between need for cognitive closure and political orientation: The mediating role of authoritarianism. *Personality and Individual Differences*, 32(4), 603-610. [https://doi.org/10.1016/S0191-8869\(01\)00062-9](https://doi.org/10.1016/S0191-8869(01)00062-9)
- Cottee, S., & Hayward, K. (2011). Terrorist (e) motives: The existential attractions of terrorism. *Studies in Conflict & Terrorism*, 34(12), 963-986.  
<https://doi.org/10.1080/1057610X.2011.621116>
- Dechesne, M. (2011). Deradicalization: not soft, but strategic. *Crime, Law and Social Change*, 55(4), 287-292. <https://doi.org/10.1007/s10611-011-9283-8>
- Doosje, B., Moghaddam, F. M., Kruglanski, A. W., de Wolf, A., Mann, L., & Feddes, A. R. (2016). Terrorism, radicalization and de-radicalization. *Current Opinion in Psychology*, 11, 79-84.  
<https://doi.org/10.1016/j.copsyc.2016.06.008>
- Efron, B., & Tibshirani, R. (1994). *An introduction to the bootstrap*. CRC Press.  
<https://doi.org/10.1201/9780429246593>
- Ferber, A. L., & Kimmel, M. S. (2008). The gendered face of terrorism. *Sociology Compass*, 2(3), 870-887. <https://doi.org/10.1111/j.1751-9020.2008.00096.x>

- Ferracuti, F. (1990). Ideology and repentance: Terrorism in Italy. In W. Reich (Ed.), *Origins of terrorism: Psychologies, ideologies, theologies, states of mind* (pp. 59-64). Cambridge University Press.
- Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1). Springer Series in Statistics. [https://doi.org/10.1007/978-0-387-21606-5\\_1](https://doi.org/10.1007/978-0-387-21606-5_1)
- Glasman, L. R., & Albarracin, D. (2006). Forming attitudes that predict future behavior: A meta-analysis of the attitude-behavior relation. *Psychological Bulletin*, *132*(5), 778-822. <https://doi.org/10.1037/0033-2909.132.5.778>
- Hafez, M., & Mullins, C. (2015). The radicalization puzzle: A theoretical synthesis of empirical approaches to homegrown extremism. *Studies in Conflict & Terrorism*, *38*(11), 958-975. <https://doi.org/10.1080/1057610X.2015.1051375>
- Hastie, T., Tibshirani, R., & Wainwright, M. (2015). *Statistical learning with sparsity: The lasso and generalizations*. CRC Press. <https://doi.org/10.1201/b18401>
- Henry, P. J., Sidanius, J., Levin, S., & Pratto, F. (2005). Social dominance orientation, authoritarianism, and support for intergroup violence between the Middle East and America. *Political Psychology*, *26*(4), 569-584. <https://doi.org/10.1111/j.1467-9221.2005.00432.x>
- Horgan, J., & Braddock, K. (2010). Rehabilitating the terrorists?: Challenges in assessing the effectiveness of de-radicalization programs. *Terrorism and Political Violence*, *22*(2), 267-291. <https://doi.org/10.1080/09546551003594748>
- Kebbell, M. R., & Porter, L. (2012). An intelligence assessment framework for identifying individuals at risk of committing acts of violent extremism against the West. *Security Journal*, *25*(3), 212-228. <https://doi.org/10.1057/sj.2011.19>
- Koehler, D. (2014). Right-wing extremist radicalization processes: The formers' perspective. *Journal EXIT-Deutschland*, *1*, 307-377.
- Koomen, W., & van der Pligt, J. (2016). *The psychology of radicalization and terrorism*. Routledge. <https://doi.org/10.4324/9781315771984>

- Krauss, S. W. (2002). Romanian authoritarianism 10 years after communism. *Personality and Social Psychology Bulletin*, 28(9), 1255-1264. <https://doi.org/10.1177/01461672022812010>
- Kruglanski, A. W., Gelfand, M. J., Bélanger, J. J., Sheveland, A., Hetiarachchi, M., & Gunaratna, R. (2014). The psychology of radicalization and deradicalization: How significance quest impacts violent extremism. *Political Psychology*, 35, 69-93.  
<https://doi.org/10.1111/pops.12163>
- LaFree, G., Jensen, M. A., James, P. A., & Safer-Lichtenstein, A. (2018). Correlates of violent political extremism in the United States. *Criminology*, 56(2), 233-268.  
<https://doi.org/10.1111/1745-9125.12169>
- McCauley, C., & Moskaleiko, S. (2008). Mechanisms of political radicalization: Pathways toward terrorism. *Terrorism and Political Violence*, 20(3), 415-433.  
<https://doi.org/10.1080/09546550802073367>
- McCauley, C., & Moskaleiko, S. (2017). Understanding political radicalization: The two-pyramids model. *American Psychologist*, 72(3), 205-216. <https://doi.org/10.1037/amp0000062>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Pfundmair, M. (2019). Ostracism promotes a terroristic mindset. *Behavioral Sciences of Terrorism and Political Aggression*, 11(2), 134-148. <https://doi.org/10.1080/19434472.2018.1443965>
- Pfundmair, M., Abmann, E., Kiver, B., Penzkofer, M., Scheuermeyer, A., Sust, L., & Schmidt, H. (2019). Pathways toward Jihadism in Western Europe: An empirical exploration of a comprehensive model of terrorist radicalization. *Terrorism and Political Violence*.  
<https://doi.org/10.1080/09546553.2019.1663828>
- Piazza, J. A. (2011). Poverty, minority economic discrimination, and domestic terrorism. *Journal of Peace Research*, 48(3), 339-353. <https://doi.org/10.1177/0022343310397404>

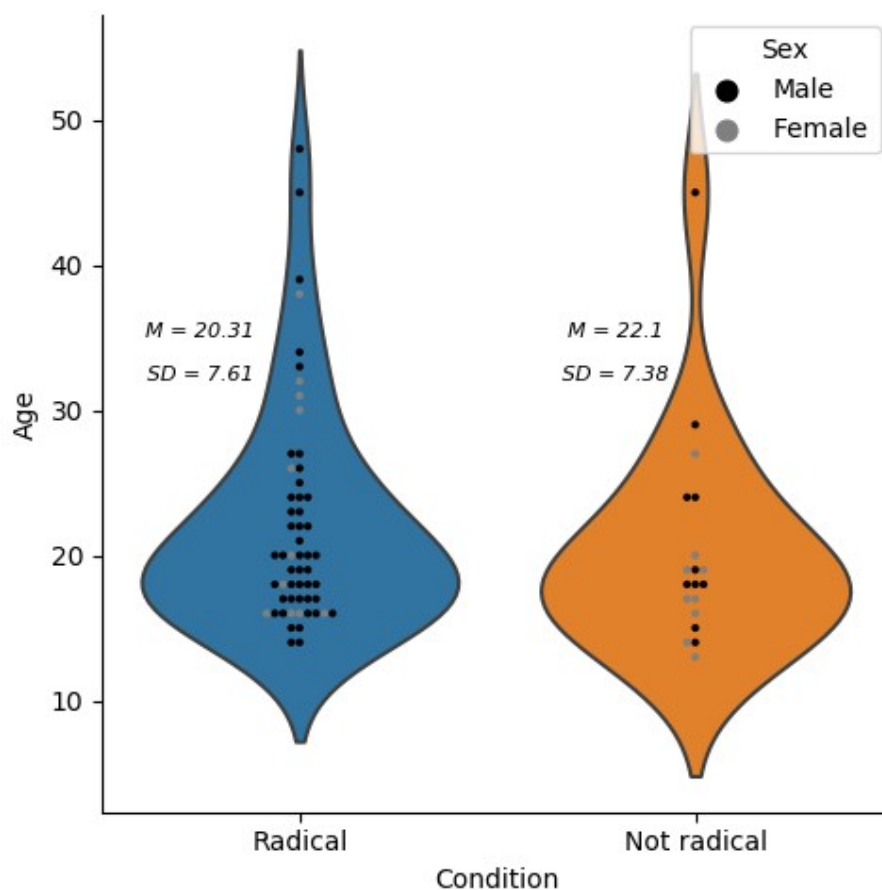
- Porter, L. E., & Kebbell, M. R. (2011). Radicalization in Australia: Examining Australia's convicted terrorists. *Psychiatry, Psychology and Law, 18*(2), 212-231.  
<https://doi.org/10.1080/13218719.2010.482953>
- Sageman, M. (2004). *Understanding terror networks*. University of Pennsylvania Press.  
<https://doi.org/10.9783/9780812206791>
- Sageman, M. (2011). *Leaderless Jihad: Terror networks in the twenty-first century*. University of Pennsylvania Press.
- Sageman, M. (2014). The stagnation in terrorism research. *Terrorism and Political Violence, 26*(4), 565-580. <https://doi.org/10.1080/09546553.2014.895649>
- Schmid, A. P. (2014). Comments on Marc Sageman's polemic "The stagnation in terrorism research". *Terrorism and Political Violence, 26*(4), 587-595.  
<https://doi.org/10.1080/09546553.2014.895651>
- Schuurman, B. (2020). Research on terrorism, 2007–2016: a review of data, methods, and authorship. *Terrorism and Political Violence, 32*(5), 1011-1026.  
<https://doi.org/10.1080/09546553.2018.1439023>
- Southers, E. (2013). *Homegrown violent extremism*. Anderson Publishing.  
<https://doi.org/10.4324/9781315721583>
- Stern, J. (2003). *Terror in the name of God: Why religious militants kill*. HarperCollins.
- Van Hiel, A., Pandelaere, M., & Duriez, B. (2004). The impact of need for closure on conservative beliefs and racism: Differential mediation by authoritarian submission and authoritarian dominance. *Personality and Social Psychology Bulletin, 30*(7), 824-837.  
<https://doi.org/10.1177/0146167204264333>
- Wiktorowicz, Q. (2005). *Radical Islam rising: Muslim extremism in the West*. Rowman & Littlefield Publishers.
- Williams, K. D. (2007). Ostracism. *Annual Review of Psychology, 58*, 425-452.  
<https://doi.org/10.1146/annurev.psych.58.110405.085641>

Wright-Neville, D., & Halafoff, A. (2010). *Terrorism and social exclusion: Misplaced risk-common security*. Edward Elgar Publishing. <https://doi.org/10.4337/9781781000700>



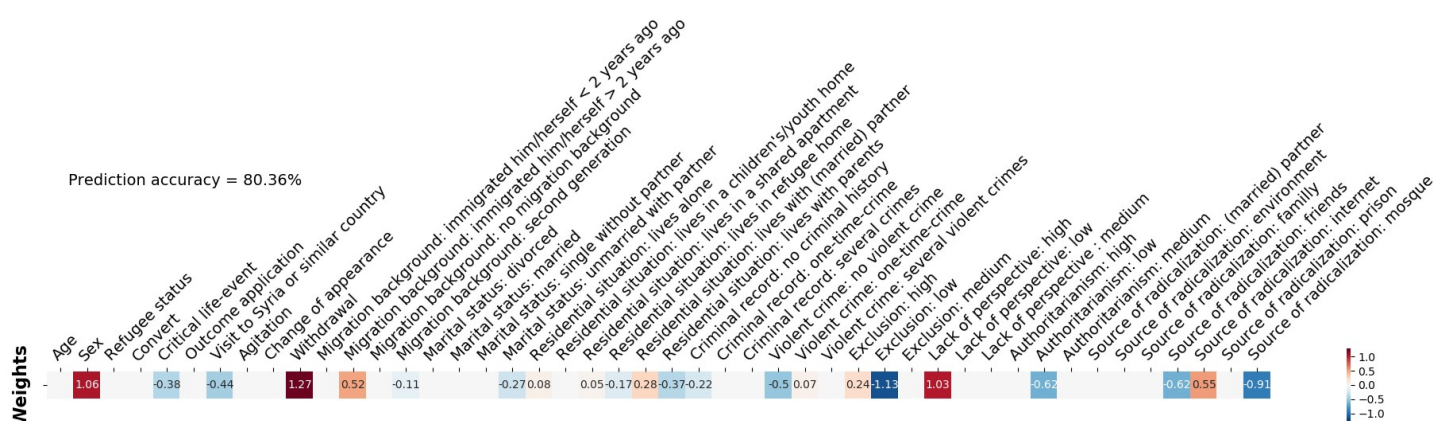
**Figure 1. Distribution of age by condition (radical or not radical) and sex (male or female)**

The blue violin plot displays the age of the radical individuals, while the orange violin plot displays the age of the individuals not radical. The width of the violins illustrates the density of the individuals' age. For instance, the shape of the first blue violin plot on the left side (skinny on each end and wide in the middle) indicates that the ages are highly concentrated around the median. The height of the violins indicates the variability (i.e., range of the individuals' age). Short violins represent a slight, while long violins represent a substantial variability. Male and female individuals are represented with black and grey dots, respectively.



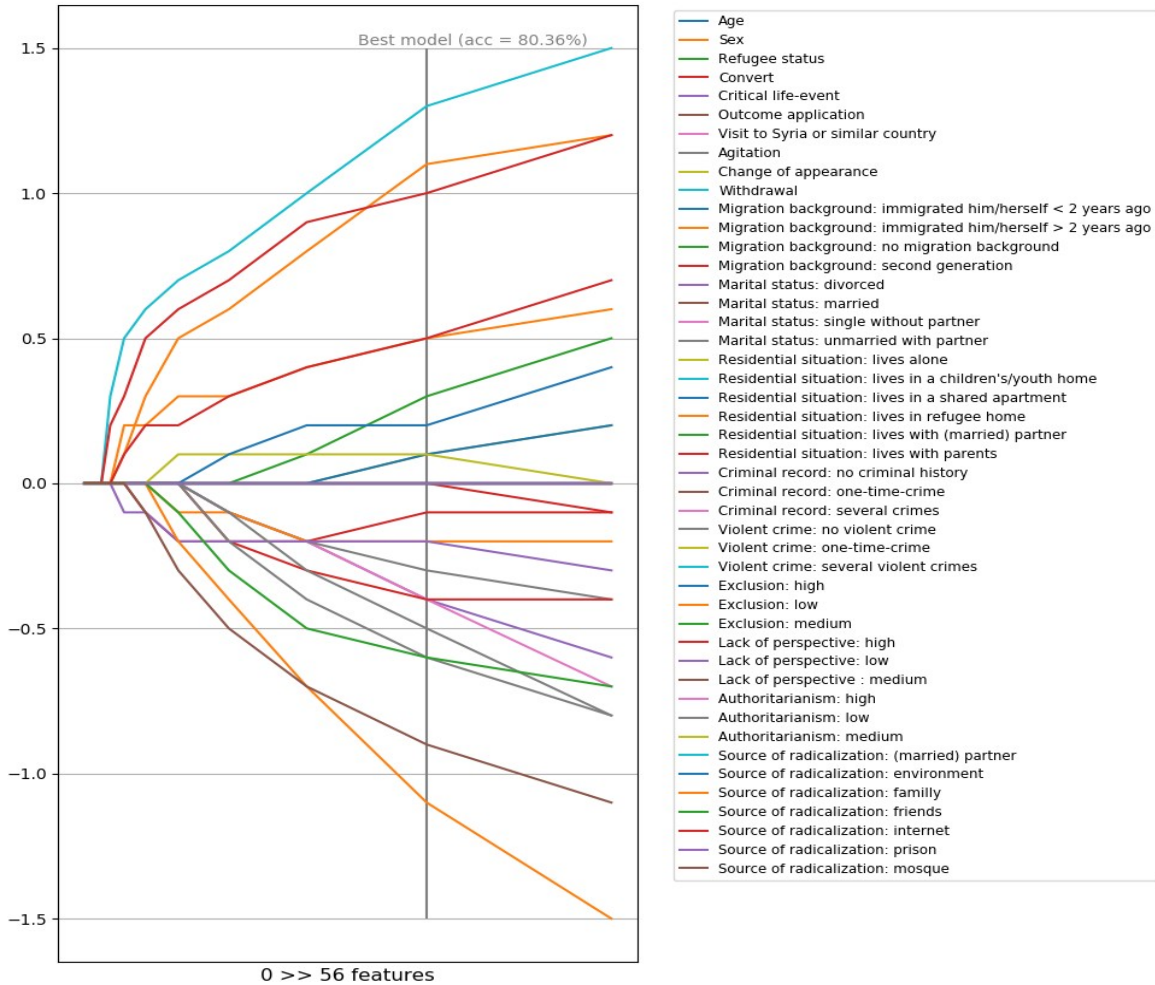
**Figure 2. Informative features for detecting radicals**

The y-axis depicts the variables included in the analysis. An empty square indicates that the variable was not included in the final model. A positive sign indicates that the feature contributed to detecting a radical. A negative sign indicates that the particular feature contributed to detecting a non-radical. For example, scoring high in the feature “source of radicalization: mosque” would tip the balance of the output toward being non-radical, while an individual having a “high lack of perspective” would tip the balance of the output toward being radical.



**Figure 3. Classifier coefficients decomposition**

Profiles of the classifier coefficients of the features are plotted on the y-axis while the decreasing parsimony constraint of this statistical model is plotted on the x-axis. The departing lines indicate changes in the selected items (i.e., the active set). The color of each line shows a particular feature. The vertical grey line shows the automatically selected model ( $\lambda = 0.75$ ).



**Figure 4. Scaling of radical individual prediction accuracy with increasing participants**

The number of participants included is plotted on the x-axis, while the y-axis displays the obtained accuracy for each specific setting (5 splits of the total amount of available data). The red line represents the average in-sample accuracy (i.e., training score), while the green line represents the average generalization performance (i.e., testing score). The green and red shadows represent the accuracy standard deviations.



**Figure 5. Estimated bootstrap confidence intervals for the prediction of radicals**

The purple circles show the estimated contribution (y-axis) of each particular feature (x-axis) to distinguishing radical versus non-radical participants using the automatically selected model in our sample. Each light blue bar indicates the bootstrapped 90% uncertainty interval at the population level.

