Research Statement

Manuel Gomez Rodriguez

June 14, 2024

My main research interest lies in the development of **human-centric machine learning**, a rapidly growing field of research that seeks to maximize the societal benefits of machine learning systems and minimize their potential harms, risks and burdens.

Since I received tenure in March 2021, my main focus has been on developing human-centric machine learning models and algorithms for evaluating, supporting and enhancing decision making, spearheaded by an ERC Starting Grant on "Human-Centric Machine Learning". My research has introduced conceptual innovations and technical breakthroughs along three different dimensions:

1. Probabilistic, causal, and game-theoretic models of decision making in specific contexts.
2. Algorithms that leverage these models to enhance decision making with provable guarantees.
3. Observational and interventional studies to evaluate such models and algorithms in practice.

Moreover, these conceptual innovations and technical breakthroughs have often uncovered previously unexplored connections between fields, as exemplified by several of my current achievements and my vision for the future.

Current Achievements

My research contributions to human-centric machine learning can be categorized in several distinctive research themes. Since each of these themes entails their own technical challenges and application domains, in what follows, I discuss each of them separately.

Human-AI Complementarity. In recent years, there has been increasing excitement about the potential of machine learning models to help human experts make more accurate predictions in a variety of application domains, including medicine, education and science. In this context, the ultimate goal is human-AI complementarity—the predictions made by the human expert who uses a machine learning model are more accurate than the predictions made by the expert or by the model on their own. However, this goal has been elusive, and it has been unclear how to design machine learning models that consistently enable human-AI complementarity. In the last years, we have pursued two lines of research to fill this gap.

In a first line of research, we have pursued the idea of algorithmic triage. Under algorithmic triage, a machine learning model does not predict all instances but instead defers some of the instances to human experts. As a result, one does not only have to find a machine learning model but also a triage policy which determines who predicts each instance. Here, one of the main challenges is that, for each potential triage policy, there is an optimal machine learning model, however, the triage policy is also something one seeks to optimize. In a sequence of papers, we have developed some of the first algorithms with theoretical guarantees to learn under algorithmic triage in regression [1], classification [2, 3] and reinforcement learning [4] settings. In these pieces of work, we have conducted observational experiments showing that, by using algorithmic triage, we can achieve human-AI complementarity, in average. However, we have also realized that algorithmic triage cannot enable human-AI complementarity at an instance level. This is because, by design, the per-instance performance is determined by either the performance of the human experts or the performance of the machine learning model.

To avoid the above limitation, in a second line of research, we have pursued the design of machine learning models that, rather than relying on algorithmic algorithmic triage, adaptively limit human experts' level of agency [5, 6]. More specifically, we have advocated for machine learning models which, rather than providing single predictions, provide sets of predictions, namely prediction sets, and ask human experts to always predict label values from these prediction sets.[1] The key rationale here is that, by using the theory of conformal prediction to construct the above prediction sets, we can precisely trade-off the probability that the ground truth label is not in the prediction set, which determines how frequently the systems will mislead human experts, and the size of the prediction set, which determines the difficulty of the prediction task the experts need to solve using the system. Further, we have developed several bandit algorithms that leverage the nested structure of the prediction sets provided by any conformal predictor and a natural counterfactual monotonicity assumption on the experts' predictions to find the conformal predictor under which experts would benefit the most from using such system very efficiently. Finally, we have conducted a large-scale human subject study that suggest that machine learning models based on prediction sets that adaptively limit experts' level of agency can enable human-AI complementarity at an instance level.

Strategic Machine Learning. As decision making is increasingly informed by machine learning models, there is an increasing pressure on the decision makers to be transparent about the decision policies, the models, and the features they use. However, individuals may be incentivized to use this knowledge to invest effort strategically in order to receive a beneficial decision. For example, in loan decisions, if a bank discloses that it will use credit card debt to decide whether it offers a loan to a customer, she may feel compelled to avoid credit card debt overall. Motivated by this observation, there has been a flurry of work in the emerging field of strategic machine learning in very recent years and we have carried out some of the early work in this field.

In a first piece of work [7], we were among the first to address the problem of finding decision policies that incentivize individuals to invest in forms of effort that increase the utility of the policy. This was in contrast with most, if not all, of the contemporary work, which had focused on the design of accurate machine learning models in the above mentioned strategic setting. One of our key ideas was the realization that individuals' investment of effort induces a change in their features and, under some technical assumptions, this change can be characterized analytically at a population level. Building upon this characterization, we have studied the hardness of the problem and identified a natural monotonicity assumption on the cost individuals pay to change features that allows for a highly effective polynomial time heuristic search algorithm to find optimal decision policies.

In a second piece of work [8], we have uncovered a previously unexplored connection between strategic machine learning and counterfactual explanations. In short, counterfactual explanations aim to help individuals subject to decisions informed by machine learning models understand what would have to change for these decisions to be beneficial ones. In this work, we have investigated how individuals may use the knowledge gained by counterfactual explanations to invest effort strategically and maximize their chances of receiving a beneficial decision. More specifically, we have developed several polynomial-time algorithms with approximation guarantees to find near-optimal policies and counterfactual explanations in such a strategic setting. Moreover, we have also shown that, by incorporating a matroid constraint into the problem formulation, one can increase the diversity of the near-optimal set of counterfactual explanations and incentivize individuals across the whole spectrum of the population to self improve.

Calibration in Machine Learning. Whenever a binary classifier is used to provide decision support, it has been widely agreed that the classifier should also provide a confidence score together with each prediction. Then, the decision maker is supposed to use this confidence score to calibrate how much to trust the prediction. In this context, the conventional wisdom is that the confidence score should be a well calibrated estimate of the probability that the predicted label matches the true label. In recent years, we have shown that, in the context of decision support, this type of confidence scores may not be as useful as thought previously.

In a first line of research, we have focused on a scenario in which human experts need to make binary decisions assisted by a classifier who makes predictions about a binary outcome of interest, and we have investigated why experts have often difficulties at developing a good sense on when to trust these predictions using the above mentioned type of confidence scores. More specifically, in a recent piece of work [9], we have shown that, for a broad class of utility functions, there exist data distributions for which an expert who makes optimal decisions would need to sometimes place more (less) trust on predictions with lower (higher) confidence scores. However, we have further shown that, if the confidence

[1] There are many systems used everyday by experts that, under normal operation, limit experts' level of agency. For example, think of a pilot who is flying a plane. There are automated, adaptive systems that prevent the pilot from taking certain actions based on the monitoring of the environment.

scores satisfy a natural human-alignment property with respect to the expert's confidence on their own predictions, there always exists an optimal decision policy under which the level of trust the expert would need to place on predictions is monotone on the confidence scores, facilitating its discoverability. Looking into the future, we are running a human subject study to investigate to what extent the above human-alignment property influences the quality of the human decisions in practice.

In a second line of work, we have focused on a scenario in which human experts need to find a subset of qualified candidates among those in a large pool of candidates and are assisted by a screening classifier that estimate the probability that a candidate is qualified. In a first piece of work [10], we have shown that, even if the estimated probabilities provided by the screening classifier are well-calibrated, in general, there always exist specific pools of candidates for which any deterministic threshold rule that shortlists candidates by thresholding the estimated probabilities will be significantly suboptimal. Then, we have developed an efficient post-processing algorithm that is able to calibrate any given classifier such that there is a deterministic threshold rule using the post-processed classifier is near-optimal in expectation across all potential pools of candidates with high-probability. In a second piece of work [11], we have shown that, even if the estimated probabilities provided by the screening classifier are well-calibrated, any deterministic threshold rule that uses such classifier may be biased against qualified candidates *within* demographic groups of interest. As a consequence, it may perpetuate historical biases against minority groups by precluding the *best* candidates within these groups—the candidates who are more likely to be qualified—to be shortlisted. Then, we have introduced an efficient post-processing algorithm to minimally modify any given calibrated classifier so that it does not suffer from this type of within-group unfairness.

Machine Learning for Counterfactual Reasoning. There is empirical evidence that humans improve their decision making skills by means of counterfactual reasoning—reasoning about what might have been had they made alternative decisions to those they actually took. However, in sequential decision making processes where multiple, dependent decisions are made sequentially over time, the number of alternatives a human may need to reason about can be intractably large, particularly if there is uncertainty on the dynamics of the environment. In the last years, we have initiated the development of machine learning models and algorithms to help humans conduct counterfactual reasoning in such settings.

In a first line of work, we have focused on Markov decision processes and developed algorithms to identify optimal sequences of alternative decisions that, in comparison with the factual decisions, would have counterfactually led to better outcomes. In a first piece of work [12], we have focused on discrete state spaces and developed an algorithm based on dynamic programming that is guaranteed to solve the problem in polynomial time. In a second piece of work [13], we have shown that, in continuous state spaces, we cannot generally expect to solve the problem in polynomial time. However, under a natural form of Lipschitz continuity of the environment's dynamics, we have developed a practical A^* algorithm that is guaranteed to return the optimal solution to the problem. In both pieces of work, using real data from the medical domain, we have shown that the alternative sequences found by these algorithms can provide valuable insights to enhance sequential decision making.

In a second line of work, we have focused on temporal point processes, a popular type of processes for modeling discrete event data in continuous time, and developed a first-of-a-kind temporal point process model that, in contrast with existing models, can be used to answer counterfactual questions about these type of processes [14]. For example, in epidemiology, assume that, during a pandemic, a government decides to implement business restrictions every time the weekly incidence—the (relative) number of new cases—is larger than certain threshold but unfortunately the incidence nevertheless spirals out of control. Our model could help the government understand retrospectively to what extent the incidence would have grown had a lower threshold been implemented.

COVID-19. Around the time I received tenure at MPI-SWS, the COVID-19 pandemic started and I quickly branched out from my current focus at that time and, together with students, postdocs and collaborators from multiple MPIs, EPFL and ETH, we worked on several projects related to COVID-19 at a furious pace. These efforts led to the development of an epidemiological model based on temporal point processes that can be used to predict the spread of epidemics at an unprecedented spatiotemporal resolution [15], a privacy-preserving and inclusive system for epidemic risk assessment and notification [16], and a group testing method based on dynamic programming that is specifically designed to use the information provided by contact tracing [17].

Previous Achievements

During my PhD and my postdoctoral work, my research focused on understanding, predicting and controlling information diffusion over the Web and social media. This led to a series of papers [18–30], which significantly advanced the state of the art in the network inference and influence maximization problems. Two of these papers [19, 24] received immediate international recognition by means of a best paper award honorable mention at KDD, the flagship conference in data mining, and an outstanding best paper award at NeurIPS, the flagship conference in machine learning. Since then, these series of papers have stimulated a large amount of follow-up work (>3,000 citations) and have been the methodological basis for two journal papers on malaria in collaboration with epidemiologists at Imperial College and others [31, 32].

After I joined MPI-SWS as a tenure-track faculty, I realized that my doctoral and postdoctoral work on network inference and influence maximization leveraged particular instances of a more general and powerful type of random processes, marked temporal point processes, which could be potentially used to design a new generation of models and algorithms to predict and optimize the functioning of social, information and networked systems. In the years that followed, I leveraged this realization to lead the design of:

(i) probabilistic models based on marked temporal point process to predict information propagation [33–35], product competition [36], opinion dynamics [37], information reliability [38–40], knowledge content [41], and spatiotemporal processes [42, 43]. In all cases, by exploiting fine grained user data, the models provide more accurate predictions than the state of the art.

(ii) a series of efficient off-line and online algorithms with provable guarantees to steer information dissemination [44–50], detect and precent the spread of misinformation [51, 52], and design spaced repetition algorithms for efficient memorization [53–55]. These algorithms exploit an alternative representation of marked temporal point processes using SDEs with jumps and establish a previously unexplored connection between optimal control of SDEs with jumps and marked temporal point processes.

At the beginning of 2020, my work on marked temporal point processes made my case for tenure at MPI-SWS. However, around the same time, I started to shift focus spearheaded by an ERC Starting Grant on "Human-Centric Machine Learning".

Vision for the Future

Looking into the future, I will further pursue the development of human-centric machine learning by means of conceptual innovations and technical breakthroughs across several distinctive research themes. In what follows, I briefly elaborate on two research themes I will particularly focus on in the next years, namely counterfactual reasoning and AI & HCI.

Counterfactual Reasoning. Had the physician initiated the antibiotic treatment for sepsis a day earlier, the patient would have recovered. Had they taken an earlier train to the airport, they would not have missed their flight. Had I clicked on the attachment of that strange email, my computer would have been hacked. These are examples of counterfactual reasoning, a type of reasoning that is epitomized by the phrase "what might have been," which implicates a juxtaposition of an imagined versus factual reality. Counterfactual reasoning is tightly connected to the way we attribute causality and responsibility, and it has been shown to play a significant role in the ability that humans have to learn from limited past experience and improve their decision making skills. Is counterfactual reasoning a human capacity that machines cannot have? Surprisingly, recent advances at the interface of causality and machine learning have demonstrated that it is possible to build machines that perform and benefit from counterfactual reasoning, in a way similarly as humans do. However, these advances have predominantly comprised machine learning systems designed to operate autonomously, without human supervision.

In the upcoming years, we will develop human-centric machine learning models and algorithms for decision support that are able to perform and benefit from counterfactual reasoning in multiple ways. For example, they will perform counterfactual reasoning about human behavior to anticipate how humans incorporate algorithmic advice into their decisions. This will enable a new generation of decision support systems that can only increase and never decrease the average quality of human decisions. Moreover, they will use the structural similarities and shared properties across different counterfactual decision making scenarios to significantly reduce their computational and data requirements. In addition, these models and algorithms will also help humans learn from their own past decisions by identifying alternative decisions that would have led to better outcomes. I have laid out this vision on an ERC Consolidator Grant proposal on "Counterfactuals in Minds and Machines", which has recently passed to the Step 2 (interview) of the evaluation.

AI & HCI. Machine learning and artificial intelligence has made it to the mainstream thanks to ChatGPT, a chatbot powered by a family of language models that has been fine-tuned to target conversational usage using a combination of supervised learning and reinforcement learning from human feedback. One of the main reasons behind its success is the intuitive user interface, which allows laypersons to refine and steer a conversation towards a desired length, format, style, level of detail, and language. On the flip side, there are increasing concerns that some of its plausible-sounding answers contain incorrect or nonsensical information and, depending on the context and the user's expertise, these *hallucinations* are difficult to detect. As a consequence, their potential to reliable enable human-AI complementarity is, in general, not very well understood.

In the upcoming years, we will develop human-centric machine learning models and algorithms that come with easy-to-use, simple interfaces that do not require humans to get a good sense on when to trust their outputs to achieve human-AI complementarity. These models and algorithms will consistently help humans with different skills and abilities solve tasks in specific contexts more efficiently and accurately. Moreover, we will go beyond the status quo in the machine learning literature, which typically relies only on benchmark datasets for evaluation, and conduct human subject studies with laypersons for evaluation, similarly as in the human computer interaction literature. Along the way, we will not only use qualitative evaluation metrics, as in most of the human computer interface literature, but also quantitative metrics and this will allow us to demonstrate that our models and algorithms, together with their interfaces, can consistently enable human-AI complementarity. Our starting point will be our recent work on machine learning models based on prediction sets [6], where we have shown that these models do not require humans to get a good sense on when to trust their predictions using a large-scale human subject study.

References

[1] A. De, P. Koley, N. Ganguly, and M. Gomez-Rodriguez. Regression under human assistance. In *AAAI*, 2020.

[2] A. De, N. Okati, A. Zarezade, and M. Gomez-Rodriguez. Classification under human assistance. In *AAAI*, 2021.

[3] N. Okati, A. De, and M. Gomez-Rodriguez. Differentiable learning under triage. In *NeurIPS*, 2021.

[4] V. B. Meresht, A. De, A. Singla, and M. Gomez-Rodriguez. Learning to switch among agents in a team via 2-layer markov decision processes. *TMLR*, 2022.

[5] E. Straitouri, L. Wang, N. Okati, and M. Gomez-Rodriguez. Improving expert predictions with conformal prediction. In *ICML*, 2023.

[6] E. Straitouri and M. Gomez-Rodriguez. Designing decision support systems using counterfactual prediction sets. In *ICML*, 2024.

[7] S. Tsirtsis, B. Tabibian, M. Khajehnejad, B. Schölkopf, A. Singla, and M. Gomez-Rodriguez. Optimal decision making under strategic behavior. *Management Science*, 2023.

[8] S. Tsirtsis and M. Gomez-Rodriguez. Decisions, counterfactual explanations and strategic behavior. In *NeurIPS*, 2020.

[9] N. L. Corvelo-Benz and M. Gomez-Rodriguez. Human-aligned calibration for ai-assisted decision making. In *NeurIPS*, 2023.

[10] L. Wang, T. Joachims, and M. Gomez-Rodriguez. Improving screening processes via calibrated subset selection. In *ICML*, 2022.

[11] N. Okati, S. Tsirtsis, and M. Gomez-Rodriguez. On the within-group fairness of screening classifiers. In *ICML*, 2023.

[12] S. Tsirtsis, A. De, and M. Gomez-Rodriguez. Counterfactual explanations in sequential decision making under uncertainty. In *NeurIPS*, 2021.

[13] S. Tsirtsis and M. Gomez-Rodriguez. Finding counterfactually optimal action sequences in continuous state spaces. In *NeurIPS*, 2023.

[14] K. Noorbakhsh and M. Gomez-Rodriguez. Counterfactual temporal point processes. In *NeurIPS*, 2022.

[15] L. Lorch, H. Kremer, W. Trouleau, S. Tsirtsis, A. Szanto, B. Schölkopf, and M. Gomez-Rodriguez. Quantifying the effects of contact tracing, testing, and containment measures in the presence of infection hotspots. *ACM TSAS*, 2022.

[16] G. Barthe, R. De Viti, P. Druschel, D. Garg, M. Gomez-Rodriguez, P. Ingo, H. Kremer, M. Lentz, L. Lorch, A. Mehta, et al. Listening to bluetooth beacons for epidemic risk mitigation. *Scientific Reports*, 2022.

[17] S. Tsirtsis, A. De, L. Lorch, and M. Gomez-Rodriguez. Pooled testing of traced contacts under superspreading dynamics. *PLOS Computational Biology*, 2022.

[18] M. Gomez-Rodriguez, D. Balduzzi, and B. Schölkopf. Uncovering the temporal dynamics of diffusion networks. In *ICML*, 2011.

[19] M. Gomez-Rodriguez, J. Leskovec, and A. Krause. Inferring Networks of Diffusion and Influence. In *KDD*, 2010.

[20] Manuel Gomez-Rodriguez, Jure Leskovec, and Andreas Krause. Inferring networks of diffusion and influence. *ACM TKDD*, 2012.

[21] M. Gomez-Rodriguez, J. Leskovec, and B. Schölkopf. Structure and Dynamics of Information Pathways in On-line Media. In *WSDM*, 2013.

[22] M. Gomez-Rodriguez and B. Schölkopf. Influence maximization in continuous time diffusion networks. In *ICML*, 2012.

[23] M. Gomez-Rodriguez and B. Schölkopf. Submodular inference of diffusion networks from multiple trees. In *ICML*, 2012.

[24] N. Du, L. Song, M. Gomez-Rodriguez, and H. Zha. Scalable influence estimation in continuous-time diffusion networks. In *NeurIPS*, 2013.

[25] M. Gomez-Rodriguez, J. Leskovec, and B. Schölkopf. Modeling information propagation with survival theory. In *ICML*, 2013.

[26] H. Daneshmand, M. Gomez-Rodriguez, L. Song, and B. Schölkopf. Estimating diffusion network structures: Recovery conditions, sample complexity & soft-thresholding algorithm. In *ICML*, 2014.

[27] Manuel Gomez-Rodriguez, Le Song, Hadi Daneshmand, and B. Schoelkopf. Estimating diffusion networks: Recovery conditions, sample complexity & soft-thresholding algorithm. *JMLR*, 2016.

[28] Manuel Gomez-Rodriguez, Le Song, Nan Du, Hongyuang Zha, and B. Schoelkopf. Influence estimation and maximization in continuous-time diffusion networks. *ACM TIST*, 2016.

[29] Nan Du, Yingyu Liang, Maria-Florina Balcan, Manuel Gomez-Rodriguez, Hongyuan Zha, and Le Song. Scalable influence maximization for multiple products in continuous-time diffusion networks. *JMLR*, 2017.

[30] M. Backes, M. Gomez-Rodriguez, P. Manoharan, and B. Surma. Reconciling privacy and utility in continuous-time diffusion networks. In *CSF*, 2017.

[31] I. Routledge, S. Lai, K. Battle, A. Ghani, M. Gomez-Rodriguez, K. Gustafson, S. Mishra, J. Proctor, A. Tatem, Z. Li, et al. Tracking progress towards malaria elimination in china: estimates of reproduction numbers and their spatiotemporal variation. *Nature Communications*, 2019.

[32] Isobel Routledge, Shengjie Lai, Katherine E Battle, Azra C Ghani, Manuel Gomez-Rodriguez, Kyle B Gustafson, Swapnil Mishra, Juliette Unwin, Joshua L Proctor, Andrew J Tatem, et al. Tracking progress towards malaria elimination in china: Individual-level estimates of transmission and its spatiotemporal variation using a diffusion network approach. *PLoS Computational Biology*, 2020.

[33] M. Farajtabar, Y. Wang, M. Gomez-Rodriguez, S. Li, H. Zha, and L. Song. Coevolve: A joint point process model for information diffusion and network co-evolution. In *NeurIPS*, 2015.

[34] M. Farajtabar, Y. Wang, M. Gomez-Rodriguez, S. Li, H. Zha, and L. Song. Coevolve: A joint point process model for information diffusion and network co-evolution. *JMLR*, 2017.

[35] C. Mavroforakis, I. Valera, and M. Gomez-Rodriguez. Modeling the dynamics of online learning activity. In *WWW*, 2017.

[36] I. Valera and M. Gomez-Rodriguez. Modeling adoption and usage of competing products. In *ICDM*, 2015.

[37] A. De, I. Valera, N. Ganguly, S. Bhattacharya, and M. Gomez-Rodriguez. Learning and forecasting opinion dynamics in social networks. In *NeurIPS*, 2016.

[38] M. Farajtabar, M. Gomez-Rodriguez, N. Du, M. Zamani, H. Zha, and L. Song. Back to the past: Source identification in diffusion networks from partially observed cascades. In *AISTATS*, 2015.

[39] Behzad Tabibian, Isabel Valera, Mehrdad Farajtabar, Le Song, Bernhard Schoelkopf, and Manuel Gomez-Rodriguez. Distilling information reliability and source trustworthiness from digital traces. In *WWW*, 2017.

[40] U. Upadhyay, A. De, and M. Gomez-Rodriguez. On the complexity of opinions and online discussions. In *WSDM*, 2019.

[41] U. Upadhyay, I. Valera, and M. Gomez-Rodriguez. Uncovering the dynamics of crowdlearning and the value of knowledge. In *WSDM*, 2017.

[42] N. Du, H. Dai, R. Trivedi, U. Upadhyay, M. Gomez-Rodriguez, and L. Song. Recurrent Marked Temporal Point Process: Embedding Event History to Vector. In *KDD*, 2016.

[43] Martin Jankowiak and Manuel Gomez-Rodriguez. Uncovering the spatiotemporal patterns of collective social activity. In *ICDM*, 2017.

[44] M. Farajtabar, N. Du, M. Gomez-Rodriguez, I. Valera, H. Zha, and L. Song. Shaping social activity by incentivizing users. In *NeurIPS*, 2014.

[45] M. Karimi, E. Tavakoli, M. Farajtabar, L. Song, and M. Gomez-Rodriguez. Smart Broadcasting: Do you want to be seen? In *KDD*, 2016.

[46] U. Upahdyay, A. De, and M. Gomez-Rodriguez. Deep reinforcement learning of marked temporal point processes. In *NeurIPS*, 2018.

[47] Ali Zarezade, Abir De, Hamid Rabiee, and Manuel Gomez-Rodriguez. Cheshire: An online algorithm for activity maximization in social networks. In *Allerton*, 2017.

[48] A. Zarezade, A. De, U. Upadhyay, H. Rabiee, and M. Gomez-Rodriguez. Steering social activity: A stochastic optimal control point of view. *JMLR*, 2018.

[49] A. Zarezade, U. Upadhyay, H. Rabiee, and M. Gomez-Rodriguez. Redqueen: An online algorithm for smart broadcasting in social networks. In *WSDM*, 2017.

[50] Khashayar Gatmiry and Manuel Gomez-Rodriguez. The network visibility problem. *ACM TOIS*, 2021.

[51] J. Kim, B. Tabibian, A. Oh, B. Schoelkopf, and M. Gomez-Rodriguez. Leveraging the crowd to detect and reduce the spread of fake news and misinformation. In *WSDM*, 2018.

[52] S. Tschiatschek, A. Singla, M. Gomez-Rodriguez, A. Merchant, and A. Krause. Fake news detection in social networks via crowd signals. In *WWW*, 2018.

[53] B. Tabibian, U. Upadhyay, A. De, A. Zarezade, B. Schoelkopf, and M. Gomez-Rodriguez. Optimizing human learning via spaced repetition optimization. *PNAS*, 2018.

[54] A. Hunziker, Y. Chen, O. Mac Aodha, M. Gomez-Rodriguez, A. Krause, P. Perona, Y. Yue, and A. Singla. Teaching multiple concepts to forgetful learners. 2019.

[55] U. Upadhyay, G. Lancashire, C. Moser, and M. Gomez-Rodriguez. Large-scale randomized experiments reveals that machine learning-based instruction helps people memorize more effectively. *npj Science of Learning*, 2021.