EPSRC
Polaris House, North Star Avenue, Swindon, Wiltshire,
United Kingdom SN2 1ET
Telephone +44 (0) 1793 444000
Web http://www.epsrc.ac.uk/

COMPLIANCE WITH THE DATA PROTECTION ACT 1998

In accordance with the Data Protection Act 1998, the personal data provided on this form will be processed by EPSRC, and may be held on computerised database and/or manual files. Further details may be found in the **guidance notes**

Standard PROPOSAL

Document Status: With Council
EPSRC Reference: EP/T027959/1

Organisation where the Grant would be held

Organisation	University of Oxford	Research Organisation Reference:	1906BL004/ED2
Division or Department	Computer Science		

Project Title [up to 150 chars]

Decentralised Arbitration

Start Date and Duration

a. Proposed start date	01 March 2020	b. Duration of the grant (months)	48

Applicants

Role	Name	Organisation	Division or Department	How many hours a week will the investigator work on the project?
Principal Investigator	Professor Paul Wilfred Goldberg	University of Oxford	Computer Science	7.5

Joint Proposals

Complete this section if more than one organisation is submitting a EPSRC proposal form for this project.

Is this part of a joint proposal ?	✔Yes No
Are you the lead RO ?	✔Yes No
Joint reference	W292340X
Total number of proposals	4

Objectives

List the main objectives of the proposed research in order of priority [up to 4000 chars]

Our primary objective is to develop a theory that will guide the design and development of novel Decentralised Arbitration (DA) systems. Currently, DA systems used in practice have been designed in a rather ad-hoc way. Such a theory should build on prior work (for example, Social Choice Theory, Mechanism Design, Algorithmic Game Theory) but address the novel features arising in this context.

We aim to guide the design of novel voting systems via mathematical analysis of models of these systems, showing them to be robust against certain attacks and vulnerabilities. We also plan to validate them in lab-based experiments.

An important objective is to understand the issues facing key stakeholders who have expressed an interest in the project topic, for example the HM Courts and Tribunals Service at the Ministry of Justice. They will operate under various constraints on the kinds of systems that would be acceptable to support Decentralised Arbitration. We seek the develop formal models of systems that can then be analysed for feasibility and performance, either via mathematical analysis or via simulation.

Design novel mechanisms for eliciting reliable responses from arbitrators, using pre-existing ideas showing how it is possible to elicit knowledge known only to a minority of arbitrators.

Design novel methods for exploiting the aspects that are not usually present in Social Choice theory, for example. This includes the possibility of payments to arbitrators, and the fact that arbitrators are often participating in a continuing rather than one-off fashion, and are managing their reputation via trust metrics.

Develop a theory of trust metrics using game-theoretic tools; introduce this to Algorithmic Game Theory community.

Pursue the applicability of these ideas in fake news detection and classification, and in other arbitration settings of interest.

We aspire to build a research community around the general topic of Decentralised Arbitration, and develop links within the UK among researchers who focus on the topic.

We seek to contribute to public engagement with these issues, by holding meetings that aim to attract representatives of organisations that face these challenges.

Summary

Describe the proposed research in simple terms in a way that could be publicised to a general audience [up to 4000 chars]. Note that this summary will be automatically published on EPSRC's website in the event that a grant is awarded.

Decentralised Arbitration refers to systems that take input from multiple participants, in order to reach a collective decision or judgement. High-profile use cases include Fake News identification, online reviewing, and distributed consensus in blockchain systems. There is also interest in designing such systems in the context of dispute resolution, where they have the potential to avoid costly legal arbitration by indicating the likely outcome of a case. Collective decision-making has of course been a challenge since the dawn of civilisation, with voting mechanisms being the main tool to achieve this. These days, novel on-line systems have the potential to exploit new features - such as trust metrics - and new forms of interaction among participants.

The project aims to develop novel models of such systems, capturing requirements of organisations with whom the team have had discussions. Our main objective is the advancement of the underlying theory.

Challenges include dealing with conflicting aims and priorities of participants, and possible bias and manipulation. Another challenge - important in the context of dispute resolution - is the need to incentivise effort to understand the case. A traditional jury is presumed to exert effort due to having the discuss the case in isolation, but in a low-cost online setting, we need novel approaches such as trust metrics. In some contexts, financial rewards may be allowable, in which case we face a problem of designing the most effective reward scheme, with attention to limited resources, and the need to avoid bias and gaming of the system.

We are generally interested in the design and analysis of "permissionless" systems having the following properties:
-- New participants may join, and existing ones may leave, at any time (and the system should be robust to arrivals/departures);
-- Truthful assessments are incentivised by means of payments and trust metrics;

-- Participants typically take part in many decisions, not just one.
There exist various systems having some of these properties, and proposals for more radical ones (e.g. Kleros) that may be used for dispute resolution, and to support so-called "smart contracts". A smart contract is an agreement that is "self-executing" in the sense that compliance with the agreement can be checked automatically, and the agreement can be carried out automatically. This may require the system to internalise certain facts about the world, a simple example being two (or more) people who bet money on the outcome of a game, in which case the game's outcome needs to be known to the system so that money can be transferred. A network of participating arbitrators (or "jurors") may provide a general-purpose means of the system internalising the relevant information. The proposed research aims to develop a theory of such systems, which will inform the development of new ones having proven performance guarantees.

This project will build on insights from Algorithmic Game Theory and Mechanism Design to develop the theory and practice of decentralised arbitration. Also relevant is work in the Economics and Management Science literature that has identified ways in which decentralised arbitrators can, for example, be used to elicit an answer to a question, even when that answer is known only to a minority of the arbitrators. The idea is to get jurors to make predictions regarding each others' predictions as well their own. We propose to extend this idea to exploit network connections amongst arbitrators, and the possibility of making payments to incentivise effort. The project proposes to analyse ideas used in on-line communities in which participants manage online content and maintain on-line reputations. These ideas may apply in scenarios in which (potentially large) networks of users are to arbitrate on an issue.

Academic Beneficiaries

Describe who will benefit from the research [up to 4000 chars].

The research is motivated by high-profile social problems and as such is of interest to diverse academic communities, discussed in more detail below. The project has close links with stakeholders outside of academia and hence one very general academic benefit will be to bring real-world issues facing these stakeholders into the research community.

Relevant research communities include:
- Algorithmic Game Theory (AGT) - the PI's immediate professional circle, whose topics include computational aspects of online systems that support multiple agents. This field overlaps with the AI community, and the analysis of algorithms community.
- Economic Theory, Applied economic modelling - members of the team in these areas also have research interests motivated by the interaction of agents with conflicting priorities
- Algorithmic Fairness - a new and growing subfield of Computer Science, whose interests intersect with the topics of the project

There are recent relevant papers with authors in systems biology, and management science. It is hoped and anticipated that the project will develop new interdisciplinary links between these communities. As a multi-site and cross-disciplinary project, it will help to develop interdisciplanry links within the UK, in particular.

The academic beneficiaries will include members of the Algorithmic Game Theory (AGT) community, and some members of the AI community. These communities study certain topics that are conceptually related to the present proposal, notably Computational Social Choice. (Social choice theory addresses the problems of combining preferences and judgements of multiple agents (e.g. via voting systems), where agents have conflicting priorities.) The topics of the present proposal have novel features that have excellent prospect for attracting interest from these researchers and leading to new questions. In the AI community, the research is relevant in particular to Multi-agent systems, and we note some related research in that area.

Academics working in Law and Legal theory are a potential beneficiary. This is because our project is motivated in part by dispute resolution and novel ways to arbitrate that avoid costly litigation (c.f. project partner HM Courts and Tribunals Service). The PI has had preliminary discussions with colleagues in the School of Law at Oxford, and will present the ideas

in this proposal in a guest lecture at a new (4th year undergraduate) course in Law and Computer Science.

Ethics in AI is emerging as a topic of interest and discussion. One challenge faced in Decentralised Arbitration is to deal with bias amongst participants in a DA system, and this introduced an ethical dimension to the topic. The design of acceptable DA systems will have to incorporate notions of fairness and transparency. There is a new ACM conference (Fairness, Accountability, and Transparency) that focuses on these issues, and is a likely venue for some of the proposed work. The proposed work aims to develop models of bias within a user community, which may constitute an appealing research direction within the area of fairness in algorithms.

There are new meetings and workshop series arising (such as "Truth and trust online" and "mechanism design for social good") whose interests have strong overlap with Decentralised Arbitration. The UK can and should develop an interdisciplinary community with strong links, and the present project will make a strong contribution to the development of this community.

Impact Summary

Impact Summary (please refer to the help for guidance on what to consider when completing this section) [up to 4000 chars]

The potential beneficiaries (outside the research community) are very diverse. Below we give more detail on three of them (project partners), associated with the topics of Fake News detection, smart contracts, and crowdsourced justice. Another general class of beneficiaries consists of online review sites, broadly understood. Well-known online review systems attempt to identify the quality of businesses, books, movies, etc. Also relevant are opinion pollsters (such as YouGov). Reviewers are self-selected, and problems include bias and moral hazard (lack of incentive to carry out due diligence in writing a review or giving scores).

Fake News - and how to recognise it - is a high-profile issue of interest to social media providers. The PI proposes to work closely with colleagues at Facebook Research (London office) who have taken an interest in the systems currently in use to address this problem. A large number of people are employed to identify possible fake news, and there are many issues of how to design the right incentives.

Several technology companies are currently developing Decentralised Arbitration (DA) systems to assist the resolution of small financial disputes. These systems use Blockchain and Smart-Contract techniques in their operation. We have already established contacts with companies like Bankex and Kleros, who expressed interest to collaborate with us within this research project. We intend to approach others who develop such systems but lack the broad academic perspective that is crucial for DA systems to make a global impact. Our findings will be disseminated across the entire industry. Firms in this industry face a variety of pressing dilemmas that require scientific and scholarly investigation. Another novel technology that will be helped by the present proposal is Smart Contracts, which various blockchain systems set out to support, and which require a theory of how the relevant facts about the world addressed by a contract may be trustlessly internalised in a blockchain system.

Dispute resolution: there have been various proposals for systems that "crowdsource justice" in an effort to resolve disputes (at least, relatively minor ones) without formal litigation. The concept is at a preliminary stage of development. Part of this project is to work through some of the the issues that are involved in designing such systems, in ways that are likely to ensure that participants can be confident of the reasonableness of the resulting decisions. In the UK there is strong motivation for such systems arising from difficulty accessing legal aid. We have had interest expressed in this project from the Ministry of Justice (HM Courts and Tribunal Service) and this will be a valuable means to stay close to the issues of importance in that domain. It should help identify obstacles and opportunities for the deployment of Decentralised Arbitration systems.

Summary of Resources Required for Project

Financial resources

Summary fund heading	Fund heading	Full economic Cost	EPSRC contribution	% EPSRC contribution
Directly Incurred	Staff	154431.31	123545.05	80
	Travel & Subsistence	26600.00	21280.00	80
	Other Costs	6500.00	5200.00	80
	Sub-total	**187531.31**	**150025.05**	
Directly Allocated	Investigators	62855.20	50284.16	80
	Estates Costs	41074.00	32859.20	80
	Other Directly Allocated	0.00	0.00	80
	Sub-total	**103929.20**	**83143.36**	
Indirect Costs	Indirect Costs	196076.00	156860.80	80
Exceptions	Staff	0.00	0.00	100
	Other Costs	0.00	0.00	100
	Sub-total	**0.00**	**0.00**	
	Total	**487536.51**	**390029.21**	

Summary of staff effort requested

	Months
Investigator	9.5
Researcher	48
Technician	0
Other	0
Visiting Researcher	0
Student	0
Total	57.5

Other Support

Details of support sought or received from any other source for this or other research in the same field.
Other support is not relevant to this application.

Staff

Directly Incurred Posts

			EFFORT ON PROJECT							
Role	**Name /Post Identifier**	**Start Date**	**Period on Project (months)**	**% of Full Time**	**Scale**	**Increment Date**	**Basic Starting Salary**	**London Allowance (£)**	**Super-annuation and NI (£)**	**Total cost on grant (£)**
Researcher	Dr PDRA 1	01/03/2020	48	100	33	01/10/2021	36914	0	13197.61	154431.31
									Total	154431.31

Applicants

Role	Name	Post will outlast project (Y/N)	Contracted working week as a % of full time work	Total number of hours to be **charged** to the grant over the duration of the grant	Average number of hours per week **charged** to the grant	Rate of Salary pool/banding	Cost estimate
Principal Investigator	Professor Paul Wilfred Goldberg	Y	100	1320	7.5	78569	62855
						Total	62855

Travel and Subsistence

Destination and purpose		Total £
Within UK	Project sites; regular meetings	6400
Within UK	Facebook Research, London; discussion meetings	1200
Outside UK	5 international conferences	11250
Outside UK	5 European Conferences	7750
	Total £	26600

Other Directly Incurred Costs

Description	Total £
Public engagement with research activities and PER training for PDRA	4000
Recruitment of PDRA	1000
Laptop for PDRA	1500
Total £	6500

Research Council Facilities

details of any proposed usage of national facilities
Research Council Facilities are not relevant to this application.

Human Participation

Would the project involve the use of human subjects?		Yes✔	No
	If yes, would equal numbers of males and females be used?	Yes	No✔
Would the project involve the use of human tissue?		Yes	No✔
Would the project involve the use of biological samples?		Yes	No✔
Would the project involve the administration of drugs, chemical agents or vaccines to humans?		Yes	No✔
Will personal information be used?		Yes✔	No
	If yes, will the information be anonymised and unlinked?	Yes✔	No
	Or will it be anonymised and linked?	Yes	No✔
	Will the research participants be identifiable?	Yes	No✔

Please provide details of any areas of substantial or moderate severity:

Our experiments will be carried out at experimental laboratories in Glasgow and Lancaster using state of the art practices and technology. In particular subjects will be receiving performance-based compensations.

While most of our research proposed here involves mathematical modeling in Game Theory and Computer Science for which no ethical issue may arise. The more minor part involving laboratory experiments will require ethics approval. The three labs mentioned above operate with standard protocols under which every experiment must be approved by the university's ethic committee. We point out, however, that the experiments we plan, which involve standard economic games, rarely pose any ethical dilemma. In most laboratories in experimental economics worldwide (including the UK) they are approved on a fast track.

We may potentially use data held by HMCTS or Kleros (see support letters). In that event, we shall implement the same strict ethics standards of HMCTS regarding privacy.

Animal Research

Would the project involve the use of vertebrate animals or other organisms covered by the Animals (Scientific Procedures) Act?	Yes	No✔

If yes, what would be the maximum severity of the procedures?	Mild or non-recovery	
	Moderate	
	Severe	
Please provide details of any areas which are Moderate or Severe:		

Animal Species

Does the proposed research involve the use of non-human primates?	Yes	✔No
Does the proposed research involve the use of dogs?	Yes	✔No
Does the proposed research involve the use of cats?	Yes	✔No
Does the proposed research involve the use of equidae?	Yes	✔No

Please select any other species of animals that are to be used in the proposed research.

Fish	Sheep
Rabbit	Rat
Amphibian	Poultry
Cow	Mouse
Reptile	Guinea Pig
Pig	Other Rodent
Bird	Other Animal

Genetic and Biological Risk

Would the project involve the production and/or use of genetically modified animals?	Yes	✔	No
If yes, will the genetic modification be used as an experimental tool, e.g., to study the function of a gene in a genetically modified organism?	Yes	✔	No
And will the research involve the release of genetically modified organisms?	Yes	✔	No
And will the research be aimed at the ultimate development of commercial or industrial genetically modified products or processes?	Yes	✔	No
Would the project involve the production and/or use of genetically modified plants?	Yes	✔	No
If yes, will the genetic modification be used as an experimental tool, e.g., to study the function of a gene in a genetically modified organism?	Yes	✔	No
And will the research involve the release of genetically modified organisms?	Yes	✔	No
And will the research be aimed at the ultimate development of commercial or industrial genetically modified products or processes?	Yes	✔	No
Would the project involve the production and/or use of genetically modified microbes?	Yes	✔	No
If yes, will the genetic modification be used as an experimental tool, e.g., to study the function of a gene in a genetically modified organism?	Yes	✔	No
And will the research involve the release of genetically modified organisms?	Yes	✔	No
And will the research be aimed at the ultimate development of commercial or industrial genetically modified products or processes?	Yes	✔	No

Approvals

Have the following necessary approvals been given by:			
The Regional Multicentre Research Ethics Committee (MREC) or Local Research Ethics Committee (LREC)?	Yes	No	Not required✔
The Human Fertilisation and Embryology Authority?	Yes	No	Not required✔
The Home Office (in relation to personal and project licences, and certificates of designation)?	Yes	No	Not required✔
The Gene Therapy Advisory Committee?	Yes	No	Not required✔
The UK Xenotransplantation Interim Regulatory Authority?	Yes	No	Not required✔
Administration of Radioactive Substances Advisory Committee (ARSAC)?	Yes	No	Not required✔
Other bodies as appropriate? Please specify.			

Other Issues

Are there any other issues of which the Council should be aware? No

Provide details of what they are and how they would be addressed [up to 1000 characters]

Project Partners: details of partners in the project and their contributions to the research. These contributions are in addition to resources identified above.

1	**Name of partner organisation**	**Division or Department**	**Name of contact**
Her Majesty's Courts		Her Majesty's Courts & Tribunals Service	Ms Kerry Greenidge

Direct contribution to project			**Indirect contribution to project**		
	Description	**Value £**		**Description**	**Value £**
cash			use of facilities/ equipment		
equipment/ materials			staff time	Discussion of issues/dialogue	5000
secondment of staff			other		
other			Sub-Total		5000
Sub-Total		0		Total Contribution	5000

2	**Name of partner organisation**	**Division or Department**	**Name of contact**
Cooperative Kleros		n/a	Dr Frederico Ast

Direct contribution to project			**Indirect contribution to project**		
	Description	**Value £**		**Description**	**Value £**
cash			use of facilities/ equipment	Joint workshop	2000
equipment/ materials			staff time	Access to internal research	15000
secondment of staff			other	Access to Data	25000

other			Sub-Total		42000
Sub-Total		0		Total Contribution	42000

3	**Name of partner organisation**	**Division or Department**		**Name of contact**	
Facebook (International)		Headquarters		Dr Eric Sodomka	
Direct contribution to project			**Indirect contribution to project**		
	Description	**Value £**		**Description**	**Value £**
cash			use of facilities/ equipment		
equipment/ materials			staff time	time spent on collaboration	40000
secondment of staff			other		
other			Sub-Total		40000
Sub-Total		0		Total Contribution	40000

Total Contribution from all Project partners	£87000

OTHER INFORMATION

Reviewers

1	**Name**	**Organisation**	**Division or Department**	**Email Address**
Professor David Manlove		University of Glasgow	School of Computing Science	david.manlove@glasgow.ac.uk

Reviewers

2	**Name**	**Organisation**	**Division or Department**	**Email Address**
Professor Rahul Savani		University of Liverpool	Computer Science	rahul.savani@liverpool.ac.uk

Reviewers

3	**Name**	**Organisation**	**Division or Department**	**Email Address**
Professor Todd Reuel Kaplan		University of Exeter	Economics	trkaplan@ex.ac.uk

Pathways to Impact: Decentralised Arbitration

The focus of this section of the proposal is the potential impact on *knowledge*, *people*, and *society*. There is potential impact on the economy if (say) project partners HMCTS or Kleros adopt ideas from the research, but the other three categories are the main ones. Plans and costs for public engagement are mentioned in the JoR.

Knowledge

We have noted earlier that Decentralised Arbitration (DA) has strong links with various subfields of Economics, AI, and Algorithmic Game Theory. They include Computational Social Choice, and Opinion dynamics on Networks. A project focused on DA thus has the potential to contribute to the exchange of ideas between these areas. There is also the AI/Law interface, mentioned in the "Research environment" paragraph of the Track record section.

The following is a more speculative example of a research area that may benefit from ideas to be developed in DA (and if the project goes ahead, we would investigate this possibility further). In the field of machine learning, *ensemble methods* refer to a general approach in which multiple models are trained on some or all elements of a data set, and in order to classify a new observation, we aggregate the predictions that are produced by the separate trained models. The study of these methods has included the analysis of alternative voting schemes that might get used to aggregate the predictions. The general study of DA has the potential to widen the scope of these methods, for example in settings where individual voters express uncertainty about their predictions.

People

The project fosters networking around the general topic of DA. Most obviously it develops new interdisciplinary links amongst the participants, but more broadly we aim to attract the involvement of further people both within and outside of the academic community, either via attending user meetings or through new collaborations or outreach activities. We will endeavour to attract participants from many backgrounds to the user workshops noted in the Workplan.

The project represents a good career development opportunity for the postdoctoral researchers, who will have the opportunity to develop links with any of the investigators on this proposal, along with project partners. With regard to the details of their skills being developed: this includes engaging with topics at the intersection of ethics and AI, which is currently a high-profile theme. We have argued for the importance of DA as a research field, having noted in this section and elsewhere in the proposal that it has very diverse applications and points of contact with other lines of research.

Society

We have identified a number of specific use cases for Decentralised Arbitration, and we plan to facilitate the transfer of theory into practice. We note in particular

Fake News detection The PI has had preliminary discussions with colleagues at Facebook Research, who are participating as a project partner. The project undertakes to pursue these interactions further, to ensure that the mathematical models we develop of arbitrators on a network, do indeed reflect what is going on in practice. We aspire to have impact on the procedures for vetting content for fake news, or on the way in which users' efforts to identify fake news are integrated into an overall judgement.

Smart contracts Several technology companies are currently developing DA systems to assist the resolution of small financial disputes. These systems use Blockchain and Smart-Contract techniques in their operation. We have already established contacts with companies like Bankex and Kleros (another project partner), who expressed interest to collaborate with us within this research project. We intend to approach others who develop such systems but lack the broad academic perspective that is crucial for DA systems to make a global impact. Smart Contracts require a theory of how the relevant facts about the world addressed by a contract may be trustlessly internalised in a blockchain system. Our interaction with Kleros gives us a useful starting-point for facilitating their usage.

HM Courts and Tribunals Service (HMCTS) We seek our engagement with project partner HMCTS to be two-sided; Eyal Winter will lead on this interaction but all the team will take part. It will involve primarily but not exclusively the behavioral aspects of our research. Depending on HMCTS's constraints and limitations, this can involve one or several of the following:

- Informing the HMCTS about our findings regarding the Nudges and soft policies that are proven effective in inducing disputers to decide in favor of arbitration or settlement rather than a formal litigation.
- Share our derived insights with HMCTS regarding juries' behavioral biases.
- The design of field experiments within HMCTS to study the effectiveness of behavioral nudges in inducing disputers to settle rather than resort to a full court procedure.
- Making use of existing HMCTS data to study the attributes of disputers and cases that make them more prone for settlement.
- Conducting discussions with HMCTS people regarding the enforcement of DA arbitration outcomes beyond reputation indices that DA companies are currently using to induce disputers to abide by the arbitration decisions.

Case for Support: Decentralised Arbitration

Paul Goldberg (PI) Eyal Winter Colin Rowat C. Sorokin

1 Research track record

This interdisciplinary project involves the design and analysis of systems that obtain a collective decision or judgement, based on input from multiple agents who may have conflicting aims and preferences. Novel features include diversity of participants, and network-based models of their interactions. The analysis draws on ideas from algorithmic game theory, social choice theory, opinion dynamics in social networks, and mechanism design. The investigating team brings together relevant expertise from Computer Science, Economics (including behavioural economics), and Management Science.

Paul W. Goldberg, PI, has an extensive track record in *Algorithmic Game Theory* (AGT), and *Theoretic Computer Science*, specifically analysis of the computational complexity of problems, and provable performance guarantees of algorithms. Highlights include an award-winning paper [4] on Nash equilibrium computation, and a new paper [7] which settles the computational complexity of two problems in the field of *fair division* that date back to the mid-1980s. He has co-authored nearly 100 papers, mostly in leading conferences and journals in the above research fields. These papers span a wide range of topics within AGT including some that are highly-relevant to the present proposal such as voting schemes [5] and opinion dynamics in social networks [6]. He is well-recognised in the field of AGT and is an associate editor of the leading CS journal *SIAM Journal on Computing*. He has served as PC chair of relevant meetings: 2012 Conference on Web and Internet Economics, 2008 Workshop on Computational Social Choice (COMSOC).

Earlier related projects: PWG was a co-investigator on an EPSRC project "Efficient Algorithms for Mechanism Design without Monetary Transfer" (joint with Liverpool and Glasgow); previously he was PI on EPSRC project "Efficient Decentralised Approaches in Algorithmic Game Theory" (a joint project of Universities of Liverpool and Warwick), and "Algorithmics of Network-sharing Games".

Eyal Winter, co-PI, is the Andrews and Elizabeth Brunner Professor of Economics at Lancaster University, specializing in Game Theory, Behavioral Economics and Decision-Making. He was awarded the Humboldt Prize for excellence in research by the German government in 2011. He is an elected council member of the Game Theory Society, and an elected fellow of the Economic Theory Society. Winter has worked extensively on multi-agent mechanism design/contracting problems and the design of incentives in teams, and on voting theory. While most of his research on this topic is theoretical some of his contributions in the area are empirical (including laboratory experiments and field data). Winter has also collaborated with CS people in research on this topic, including Moshe Babaioff, Michal Feldman and Noam Nisan. Winter has published in leading journals in Economics and Social Sciences including *Econometrica*, *AER*, *ReStud*, *APSR*, *JET*, *Management Science*, *Psy Science*. His press essays appeared in *Time Magazine*, *Forbes*, *Los Angeles Times*, *Washington Post*, *Guardian*, *the Independent*, *Jewish Chronicle*, *The Conversation*, *Haaretz* and more. He has advised governments, corporations and NGOs. Relevant papers include [1, 9–11, 13].

Colin Rowat, co-I, has worked for over a decade at the interface of theoretical economics and computer science. He has held an EPSRC grant to apply *formal methods* from computer science to problems in economics. This work culminated in a sound implementation of a combinatorial auction design [3]: properties were proved on the formal specification, which was then compiled to executable code, providing provable performance guarantees that the implemented version maintained the specified properties. He has been a guest lecturer at the Initiative in Computational Economics, co-organised a formal methods track at the AISB (Society for the Study of Artificial Intelligence and Simulation of Behaviour), and guest edited a special issue of *Mathematics in Computer Science*. His most recent paper is in 'explainable AI', developing a measure of features' importance that incorporates structural information such as causal information [8]. Since 2016, he has given public

engagement lectures as part of the ESRC's Festival of Social Science, focusing on big data and AI; he has a TEDx talk on formal methods.

Constantine Sorokin, co-I, is a Lecturer in Economics, Glasgow University, Adam Smith Business School and a senior research fellow at the Game theory laboratory, Higher School of Economics, St. Petersburg, Russia. His fields of expertise are Game Theory, Political Economy, Mechanism Design, Conflict Resolution. He has ongoing collaboration with Eyal Winter. He has worked on theory (stochastic voting models, as well as information design problems in auctions and conflict escalation model, e.g. [12, 14]) and has participated in a number of applied projects; a recent relevant one for the present proposal is: "Decentralized Arbitration as a part of Digital Deal platform" (Contract with Bankex, Moscow).

Research Environment The PDRA at Oxford will be based with the PI in the Algorithms and Complexity Theory research group, which has wide-ranging expertise in the analysis of algorithms and their provable performance guarantees. There is relevant expertise in the department's AI and Machine Learning research theme (in particular, in Computational Social Choice). The PI also has close links with members of the Economics department at Oxford ([2] is a recent joint paper).

The PI is in touch with colleagues in the Law school at Oxford, who have an interest in the Law/AI interface. The PI will present the ideas in this proposal in a session of a new undergraduate course on Law and Computer Science `http://www.cs.ox.ac.uk/teaching/courses/2019-2020/LawandCS/`.

The lead research organisation (Dept. of Computer Science, Oxford) has hosted a couple of more distantly-related projects motivated by the Fake News problem, which is a use case of Decentralised Arbitration. One of these is the "digital wildfire" project, based partly at Oxford: `http://www.digitalwildfire.org/`; Its focus differs in that the "digital wildfire" project addresses regulation of social media, as a means to control fake news, while here we study incentives and decentralised mechanisms (how to help the identification of fake news, and real news). Also, our use cases go beyond the "fake news" topic. But researchers on the "digital wildfire" are a useful resource. The other one, FakeNewsRank (of Thomas Lukasiewicz) aims at automatic Fake News identification based on its textual content.

Selected Publications

[1] M. Babaioff and E. Winter. Contract complexity. In *15th ACM Conference on Electronic Commerce (ACM-EC)*, 2014.

[2] E. Baldwin, P. W. Goldberg, P. Klemperer, and E. Lock. Solving strong-substitutes product-mix auctions. *CoRR*, abs/1909.07313, 2019.

[3] M. B. Caminati, M. Kerber, C. Lange, and C. Rowat. Sound auction specification and implementation. In *16th ACM Conference on Economics and Computation, ACM-EC*, pages 547–564, 2015.

[4] C. Daskalakis, P. W. Goldberg, and C. H. Papadimitriou. The complexity of computing a Nash equilibrium. *SIAM Journal on Computing*, 39(4):195–259, 2009.

[5] E. Elkind, L. A. Goldberg, P. W. Goldberg, and M. J. Wooldridge. On the computational complexity of weighted voting games. *Annals of Mathematics and Artificial Intelligence*, 56(2):109–131, 2009.

[6] D. Ferraioli, P. W. Goldberg, and C. Ventre. Decentralized dynamics for finite opinion games. *Theoretical Computer Science*, 648:96–115, 2016.

[7] A. Filos-Ratsikas and P. W. Goldberg. The complexity of splitting necklaces and bisecting ham sandwiches. In *51st Annual ACM Symposium on the Theory of Computing*, pages 638–649, 2019.

[8] C. Frye, I. Feige, and C. Rowat. Asymmetric shapley values: incorporating causal knowledge into model-agnostic explainability. *CoRR*, abs/1910.06358, 2019.

[9] A. Gershkov and E. Winter. Formal vs. informal monitoring in teams. *American Economic Journal: Microeconomics*, 7(2):27–44, 2016.

[10] J. Kagel, H. Sung, and E. Winter. Veto power in committees: An experimental study experimental economics. *Experimental Economics*, 13(2):167–188, 2010.

[11] E. Klor, S. Kube, E. Winter, and R. Zultan. Can higher rewards lead to less effort? incentive reversal in teams. *Journal of economic behavior and organization*, 97:72–83, 2014.

[12] C. Sorokin and A. Zakharov. Vote-motivated candidates. *Journal of Economic Theory*, 176:232–254, 2018.

[13] E. Winter. Incentives and discrimination. *American Economic Review*, 94(3):764–773, 2004.

[14] A. V. Zakharov and C. S. Sorokin. Policy convergence in a two-candidate probabilistic voting model. *Social Choice and Welfare*, 43(2):429–446, Aug 2014.

2 Proposed Research and Its Context

Executive Summary: Decentralised Arbitration

The project aims to model and analyse networks of arbitrators (or jurors) who reach collective decisions. Use cases include “fake news” identification, recommender systems, legal dispute resolution, and (in the context of blockchain) agreement on transaction history. Our project partners indicate the diversity of use cases. The research will apply ideas from AI, social choice theory, and pricing/auction theory. We will develop new models of these settings, with the aim of providing guidance on the design of the right incentives for arbitrators to cooperate to identify the right answer. We also propose to conduct game-theoretic lab experiments. This interdisciplinary project combines AI, economic theory, and legal theory.

Keywords: Algorithmic Game Theory, AI, crowdsourcing, provable performance guarantee

2.1 Background

Decentralised Arbitration (DA) refers to systems that support a network of participants, whose task is to assess the correctness/incorrectness of various statements about the outside world. The ones of interest to the present proposal have certain novel features:

1. New participants may join, and existing ones may leave, at any time (and the system should be robust to arrivals/departures); systems like this are said to be “permissionless”, contrasting sharply with traditional juries and committees;

2. Truthful assessments — and the effort required to identify a good assessment — are incentivised by means of payments and trust metrics (or reputation systems);

3. Participants may take part in many decisions over time (trust metrics reflect track record).

The proposed research addresses a theory of such systems, which will inform the development of new ones having proven performance guarantees and improved performance.

Some of the general ideas associated with DA can be observed at work in well-known internet platforms such as Wikipedia, Reddit, and Stack Exchange. These platforms use mechanisms for participants to approve each others’ contributions, and have gained trust amongst users. Other use cases are mentioned in this proposal; an emerging one is the need for distributed ledger systems to maintain agreement on transactions, without requiring all participants to “remember everything”. The trust metrics used in the above web sites have been designed in an ad-hoc manner. This motivates the development of the theory of how these mechanisms can/should work.

Pre-existing lines of research have led to tools and techniques that feed in to the present project.

1. **Social networks and opinion dynamics** This field studies networks of agents who influence their neighbours. Agents are assumed to benefit from agreeing with their neighbours. Research issues include the identification of influential agents within a given network [21], and prediction of what consensus may be reached via opinion dynamics [15] [6].

   A recent interdisciplinary paper [28] investigates experimentally how crowd-sourced decisions can be manipulated by an unscrupulous network designer. We are interested in analysing a mathematical model of this scenario, with a view to further understanding the prizes and pitfalls of using arbitrators on a network.

2. **Mechanism design** from Economics, which includes matching systems as well as auctions and pricing mechanisms, is a major theme in Algorithmic Game Theory (AGT).

   Mechanism design without money is an important subtopic pursued in AGT and AI; relevant real-world problems include aggregation of reviews, e.g. code review in large software companies. The present proposal aims to extend some of these ideas to systems with limited financial incentives and trust metrics, and thus go further in the range of use cases envisaged. Another relevant topic is “information elicitation without verification”, e.g. [23].

3. **Computational social choice**, which has been studied intensively in AI during the past 10 years; it studies how to aggregate preferences of an electorate, usually without the use of payments. Here we introduce novel aspects discussed further in Section 2.5.1.

4. **Legal theory**, e.g. contract theory, which is beginning to attract attention from the AGT community [18]. Generally, the AI/Law interface is attracting increasing interest [22].

Another general inspiration is the celebrated PageRank algorithm (and related algorithms such as HITS) that rank web pages in a way that may be regarded as a decentralised mechanism: links are regarded as endorsements of web pages, and the overall rank of a web page arises from its peers, and *not* from external assessment of its contents.

Decentralised arbitration (DA) refers to settings where a single system takes input from multiple contributors (thought of as jurors); it is a main interest of project partner Kleros. Crucially, arbitration may be facilitated by payments or other incentives, thus combining elements of algorithmic pricing and social choice. There is a need for incentive systems designed via game-theoretic analysis of these novel settings. Note that in the settings of interest, there is a "right answer" and we aim at consensus in support of that right answer. That contrasts with settings of social choice where the objective is an outcome that is as much as possible in line with the preferences of the voters (and there is no "correct" outcome). *The above-listed topics clearly relate to this, but DA introduces new issues.* DA introduces issues of design of incentives, legal/regulatory aspects, preferences/selfishness of participants, how to persuade contracting sides to use a DA system, and arbitration as a repeated or on-line process, rather than one-off.

The present proposal is strongly motivated by real-world problems, together with the rich theories that have been developed by the above research topics, and their potential to be applied. It looks timely to exploit the big ideas that have been developed, in the context of novel problems.

2.2 National Importance

In the UK, a high-profile issue is access to legal aid, and its affordability. It is hoped, amongst other things, that this methodology may help arbitrate minor disputes and avoid litigation, at a time when legal aid has become hard to access, for people who cannot afford legal representation. See the support letter from HM Courts and Tribunals; this project will be an opportunity to engage with this issue.

Another use case of the present proposal is fake news detection and control, of interest to social media providers. The public interest of this use case is clear. The PI's colleagues at Facebook Research (London office) will advise on the issues that are specific to this use case.

EPSRC portfolio: EPSRC has identified AI technologies to grow as a proportion of its portfolio, and "AI and data" have been named as one of four grand challenges of the UK Govt's industrial strategy. EPSRC supports AI at the doctoral level, in the form of 16 new CDTs in AI.

2.3 Academic Impact

This project develops interdisciplinary links between Computer Science and Economics that are important for this and other endeavours at this interface (such as online markets, and game-theoretic modelling of robot interaction).

The project raises new and interesting questions at the AI-and-ethics interface, currently something of a hot topic. In particular, how to address the issue of bias and discrimination among arbitrators.

Timeliness: new and relevant meetings are emerging such as the Conference for Truth and Trust Online (TTO) `https://truthandtrustonline.com/`, and the Mechanism Design for Social Good workshop (MD4SG) `http://www.md4sg.com/`. The issue of addressing jurors' biases (see WP1) has a tie-in with the new annual ACM conference on Fairness, Accountability and Transparency `https://fatconference.org/`, and we aspire to introduce this new line of research to that community. There is a direct connection with AGT since both are concerned with the interactions of multiple agents, and their social and individual welfare.

2.4 Research Hypotheses and Objectives

Main hypothesis: there exist novel and unexplored ways of crowdsourcing accurate judgements, that are more reliable than a naïve majority-vote "ask the audience" approach. The work of Prelec [25,26] suggests a way forward: a carefully-designed voting mechanism can elicit a correct answer that is only known to a (possibly small) minority of the electorate. A further hypothesis is that mathematically rigorous analysis is needed in order to assure users of the reliability of DA systems. Ideas from *prediction markets* [30] may also be valuable.

Main objective: develop a theory of what can and cannot be achieved, in the above-described setting of DA. What assumptions do we need to make about participants in a DA system, in order to get reliable arbitration? For example, it looks helpful to assume that they value a track-record of accurate judgements, but we need to formalise this idea, in order to apply it rigorously (see [19]). Based on this modelling, we aim to identify optimal mechanisms for the task of arbitration. Having developed the theory, we will investigate applications (c.f. our project partners).

2.5 Programme and Methodology

We divide the programme into 5 Work Packages (WPs). All the team work equally on WP1; WP2 is led by PWG and EW; WP3 is led by CR; WP4 is led by CR/PWG; WP5 is led by EW and CS.

2.5.1 WP1: General theory of decentralised arbitration

Develop original theory where pre-existing work (e.g. mechanism design) does not satisfactorily address problems faced by practitioners involved in leading DA use cases. (Subsequently, WP5 will experimentally validate the theory.)

- **Voting rules.** The field of Computational Social Choice has introduced the study of numerous novel voting rules, addressing aspects of computational complexity, manipulability, tractable classes of preference functions, and other issues. We intend to build on these ideas but in the context of novel issues arising in the present domain:
  - uncertainty expressed by voters (arbitrators), and uncertainty in the outcome of the vote.
  - emphasis on search for "ground truth" rather than consensus.
  - Voting may be a repeated game in which trust is being built-up by voters (see WP4).
  - Voting rules should typically not treat all voters equally, raising new questions about how to bias towards "reliable" voters while attracting new participants.
- **Incentive design.** How should jurors be incentivised to exert effort towards their voting decision? In some DA systems it is allowed to introduce financial incentives. Should a juror receive a higher compensation for his/her services if his/her cast vote coincides with that of the majority? Should jurors pass a preliminary quiz to check their acquaintance with the case before casting their vote? What kind of behavioral nudges will increase jurors' motivation to engage themselves with the underlying case?
- **Fairness.** How should jurors' biases, affected by irrelevant attributes of disputers (eg. ethnicity, gender, firms vs. customers) be minimised or taken into account when aggregating votes? (c.f. [24]) This issue becomes more complicated when arbitrators are assumed to lie on a network (WP2).
- **Uncertainty** of a collective judgement is a novel feature motivated by DA for Fake News detection, in which an uncertain judgement can be referred for further scrutiny. How should probabilistic (individual) judgements be aggregated into a probabilistic collective one? Even without probabilities associated with individual judgements, it may be appropriate to explain disagreement between them in terms of collective uncertainty.

Other general issues include the following:

- How should jurors' services be organised dynamically? Should jurors be incentivised to specialise in specific fields in which they will carry out judgements?

- How can disputers be incentivised to use a DA system rather than seeking justice through standard courts? What kind of soft policy and behavioral nudges can help facilitate the transition from classical courts to DA? How should debate (if any) amongst jurors be structured?

2.5.2 WP2: Decentralized Arbitration through social networks

We give a detailed example to illustrate the kind of problem of interest (the proposal aims to develop it further). Subsequently we discuss related problems at a more general level.

One reasonable option for recruiting arbitrators is doing it through online networks. Indeed, this mode of recruiting is particularly relevant to the problem of fake news identification, of interest to Facebook and others. How to control fake news raises ethical issues [20]. Co-PI Eyal Winter has considered the following general problem which involves the design of optimal mechanisms for the initiation of actions in social networks. In our specific context here the action can either be agreeing to serve as arbitrator or initiating arbitrators to exert the necessary effort to produce a quality decision.

There are N individuals (agents) each of whom decides whether to take some action. Agents are related through a social network, which is represented by a symmetric matrix (graph) G, where if i and j are connected we let $G_{ij} = G_{ji} = 1$ and $G_{ij} = 0$ otherwise. We denote the number of neighbors of i in a network by d_i, and assume that $d_i > 0$. The decision whether to take the action that the principal wishes to induce (joining as an arbitrator on a particular task) yield binary strategies $s_i \in \{0, 1\}$.

All players are exposed to social influence by their network neighbors. We assume this social influence has the following functional form $in_i(s) = f(d_i)d_i^*$ where $d_i^* = \{j : G_{ij} = 1 \text{ and } s_j = 1\}$ denotes the number of i's neighbours who have taken the action, and $f(d_i)$ is a dilution function (decreasing in d_i). Players i's utility function over strategy profiles is now given by

$$u_i(s) = \alpha_i + in_i(s) + t_i$$

where α_i is the idiosyncratic taste parameter regarding the action, $in_i(s)$ is the social influence and t_i is an external influence (transfer) generated by the principal. For any given vector of transfer t agents participate in a game with equilibrium behaviour determining whether they act or not.

As mentioned, the principal wishes to induce agents to act. An optimal inducing mechanism is a vector of transfers t satisfying the following two conditions. (1) The underlying game has a unique equilibrium with all agents acting, and (2) there is no other vector of transfer t' satisfying (1) with a lower total transfer. It has been shown that agents who are more connected receive preferential treatment in the optimal mechanism.

Optimality of networks. When recruiting arbitrators (or when inducing them to exert effort), the principal can also control the structure of the network. This can be done by controlling the information flow between nodes, or by selecting specific nodes (users) out of the larger network of Facebook friendships. This has the potential to build teams of arbitrators having complementary skills and track-records, but also raises the spectre of "information gerrymandering" [28].

Optimal network design. Because of the dilution effect networks will typically not be complete. The initial indication is that the optimal networks must be nodes is partitioned to two sets H (for hubs) and P (for periphery). Nodes in H are connected to all other nodes, while nodes in P are connected only to those in H. Furthermore hub nodes earn higher transfers in the optimal mechanisms relative to those in the periphery.

Modelling partial knowledge of network structure Informally, the notion of a network of arbitrators refers to the assumption that some pairs of arbitrators may influence each other (and their payoffs) more strongly than others. In most work in social choice theory, the electorate is unstructured (there is no network). At an opposite extreme, the network is a graph whose vertices are the arbitrators, and whose structure is known at all times. We may revisit these questions for settings of partial knowledge of the network; details of models to be discussed with project partners.

2.5.3 WP3: Smart contracts

A smart contract refers to any agreement that is automated, often in the context of blockchain systems. Systems that support smart contracts need to internalise certain facts about the real world: for

a simple example, suppose two users bet on who wins a football game, if the system can internalise the outcome of that game, then the bet can be settled automatically. Clearly one could, in an ad-hoc manner, identify an authoritative source on the issue of who won the game of interest. The challenge is to devise a generic mechanism that elicits that fact (of who won the game) from other participants in the system in a reliable way.

The online arbitration service Kleros intends to support smart contracts, for details see the white paper `https://kleros.io/assets/whitepaper.pdf`. We note that the incentive systems described therein, are somewhat ad-hoc and lack proven performance guarantees.

Challenges involving analysis of smart contracts There are recent papers that include some mathematical analysis but have gaps in the analysis. Adler et al. [16] proposes a system that supports smart contracts by eliciting relevant external data from participants. There is some game-theoretic analysis of properties of the system, but with some gaps. For example, no guarantee is established regarding the non-existence of undesirable Nash equilibria. Indeed, it would be preferable to rule out alternative solution concepts e.g. *correlated* equilibria that do not achieve the system's objectives. Wang and Malluhi [29] discuss implementation of a particular type of smart contract in blockchain, and give an impossibility result; this raises many interesting questions regarding what can be achieved, in this domain.

2.5.4 WP4: Towards an economic theory of trust metrics

In law, jurors serve in a single case in a self-contained manner: there is no notion of "track-record" that is applied to any juror — a key principle is that all are equal. By contrast, in the systems we envisage as use cases, participants develop reputation over time, represented via trust metrics (for example recommendations and endorsements from peers in the network). Existing systems exploit the phenomenon that people seek to develop this kind of recognition, and it is clearly valuable to exploit it. The general research issue is the principled design of trust metrics. How to measure trust, and to what extent do we want to give extra voting power to highly-trusted arbitrators? Trust and reliability should be exploited, but is there a risk we go too far and allow a single arbitrator to impose a bad judgement? This ties in with the issue we mentioned earlier, about representing uncertainty in judgements. Another novel aspect: in reputation relative rather than absolute scores are important; game-theory models known as *contests* look appropriate.

Trust has been measured via experimental games whose Nash equilibria differ from Pareto optimal behaviour (see the "Games" section of the Wikipedia page on trust metrics). Development of this line of work would be a novel and original contribution to the research topics of Algorithmic Game Theory.

2.5.5 WP5: Lab experiments

To test the validity of our incentive mechanisms for arbitrators derived from our game-theoretic analysis we plan to study them via laboratory experiments. Lab experiments are essential because they will allow us to fully control the arbitrators' strategic environment. In a real DA environment we will never be able to verify (even long after the arbitrators' decisions) which one is "correct". Hence, we will not be able to fully study how different incentive schemes affect the quality of arbitrators' decision-making. In contrast the lab allows us to address problems to the virtual juries for which the correct answers are known, which will allow us to monitor jurors' performances with respect to different incentive schemes. We next describe the general framework of these experiments. Their specific design is part of the research project itself and will depend on the theoretical results and insights we obtain, along with feedback from project partners.

One set of experiments deals primarily with the incentives to exert effort towards making an educated decision in the DA process. Another set studies voters' potential biases (such as conformism/herding or prejudice) and the way they can be diminished.

For the first set of experiments we intend to use tasks/decisions that have clear success/failure outcomes but require a certain degree of effort to perform successfully (answering questions that require the careful reading of a short text or the performing of some simple calculations). We shall test both simultaneous voting as well as sequential voting. In the latter, voters will receive signals

regarding the success of their predecessors' decision before they decide whether to invest effort in their own decision. Within this framework we will test several mechanisms that map voters' individual decisions to a collective outcome and assign payoffs to voters. These mechanisms will differ in terms of the contracts' complexity. Some will be simple and require that the payment depends only on the success of the collective decision and some will be more complex and allow payoffs to depend on a more detailed description of individuals' decisions.

Our second set of experiments will attempt to investigate potential behavioral biases in DA environments. We shall ask questions such as: Do voters tend to conform to the decisions of their peers? Do they exhibit group favoritism (e.g. rule in favour of a student if they themselves are students)? Do voters' political orientation affect their choices (e.g. in ruling on a dispute between a firm and a customer)? In these experiments we will generate virtual scenarios of disputes on which subjects will have to vote. While subjects will know that the scenarios are virtual the DA game will be real. Subjects will vote on decisions and payoffs will be allocated according to a pre-specified mechanism. Answers to the questions posed above will be derived by means of presenting different scenarios, collecting information about the subjects through questionnaires and manipulating the information available to voters.

Our experiments will be carried out at an experimental lab in Lancaster using state of the art practices and technology. Subjects will receive performance-based compensations.

In contrast with the other WPs, this part involving laboratory experiments will require ethics approval. The lab mentioned above operates with standard protocols under which every experiment must be approved by the university's ethic committee. We point out, however, that the experiments we plan, which involve standard economic games, rarely pose any ethical dilemma. In most laboratories in experimental economics worldwide (including the UK) they are approved on a fast track.

References

[15] R. Abebe, J. M. Kleinberg, D. C. Parkes, and C. E. Tsourakakis. Opinion dynamics with varying susceptibility to persuasion. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD*, pages 1089–1098, 2018.

[16] J. Adler, R. Berryhill, A. G. Veneris, Z. Poulos, N. Veira, and A. Kastania. ASTRAEA: A decentralized blockchain oracle. In *IEEE iThings/GreenCom/CPSCom/SmartData*, pages 1145–1152, 2018.

[17] L. M. Ausubel and P. Milgrom. The lovely but lonely Vickrey auction. In *Combinatorial Auctions, chapter 1*. MIT Press, 2006.

[18] P. Duetting, T. Roughgarden, and I. Talgam-Cohen. Simple versus optimal contracts. In *20th ACM Conference on Economics and Computation*, 2019.

[19] E. J. Friedman and P. Resnick. The social cost of cheap pseudonyms. *Journal of Economics & Management Strategy*, 10(2):173–199, 2001.

[20] B. J. Grosz, D. G. Grant, K. Vredenburgh, J. Behrends, L. Hu, A. Simmons, and J. Waldo. Embedded ethics: integrating ethics across CS education. *Commun. ACM*, 62(8):54–61, 2019.

[21] D. Kempe, J. M. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 137–146, 2003.

[22] L. Kugler. AI Judges and Juries. *Commun. ACM*, Dec. 2018.

[23] Y. Liu and Y. Chen. Machine-learning aided peer prediction. In *18th ACM Conference on Economics and Computation, EC*, pages 63–80, 2017.

[24] S. G. Mayson. Bias in, bias out. *128 Yale Law Journal 2218 (2019), University of Georgia School of Law Legal Studies Research Paper No. 2018-35*, 2018.

[25] J. McCoy and D. Prelec. A statistical model for aggregating judgments by incorporating peer predictions. *Arxiv preprint*, 2017.

[26] D. Prelec. A Bayesian truth serum for subjective data. *Science*, 306(5695):462–466, 2004.

[27] M. H. Rothkopf. Thirteen reasons why the Vickrey-Clarke-Groves process is not practical. *Operations Research*, 55(2):191–197, 2007.

[28] A. J. Stewart, M. Mosleh, M. Diakonova, A. A. Arechar, D. G. Rand, and J. B. Plotkin. Information gerrymandering and undemocratic decisions. *Nature*, 573:117–121, 2019.

[29] Y. Wang and Q. M. Malluhi. The limit of blockchains: Infeasibility of a smart obama-trump contract. *Commun. ACM*, 62(5):64–69, 2019.

[30] J. Wolfers and E. Zitzewitz. Prediction markets. *Journal of economic perspectives*, 18(2):107–126, 2004.

Justification of Resources: Decentralised Arbitration

Staff Costs

Investigator time. Each investigator commits 20% of their time (one day per week) over the course of the project. This will allow the investigators to engage fully with the technical challenges and the interdisciplinary collaboration which is a key feature of the project, also to meet up regularly.

Roles of the investigators All investigators will be involved in the analysis of novel mathematical models of DA. EW, CS will conduct the lab experiments later in the project, with assistance from PDRA 2. PWG and PDRA 1 will interact with colleagues at Facebook on issues arising in Fake News identification. EW is the main contact for colleagues in Kleros and HM Courts and Tribunals Service.

Postdoctoral research associates (PDRAs).

- PDRA 1 - 100% for a 3 year term, based with the PI at Oxford. PDRA 1 will contribute to the development of the general theory of DA, work on the details of the proofs and analysis, and present results at conferences and workshops.
- PDRA 2 - 100% for a 2.5 year term, based with EW at Lancaster. PDRA 2 will carry out much of the planned experimental work: this person will help with the recruitment of subjects to the experiments, administrate the experiments, and will use statistical software packages to analyze the data. PDRA2 will also contribute to analysis, writing-up, and presentation.

Both PDRAs are costed at grade 7, commensurate with the experience required for this project and a total of £251,615 is requested between the two sites.

A total of £2,000 (split between Oxford and Lancaster) is requested for recruitment costs.

Non staff Costs

Workshops, research visits, and travel

The research topic of this proposal is connected with the following international conferences, which mostly take place in the USA, sometimes in Europe. Presentation of the work at these prestigious meetings is necessary for publicising it and obtaining feedback into research directions.

- ACM Conference on Fairness, Accountability, and Transparency
- ACM Conference on Economics and Computation

It is likely that the research will also contribute to general AI conferences which again are mostly are held in the USA. The main ones are AAAI and IJCAI.

We request funds for each investigator to attend smaller-scale meetings for promoting the research topic and gaining input from additional prospective users. A number of new and relevant workshops and conferences are appearing in the UK and elsewhere (indicating the timeliness of the topic). For example the investigators have been invited to attend a one-day conference “When Arbitration Meets Blockchain: the Birth of Decentralized Justice” at the University of Leicester Law School (in Jan. 2020). Another example is the new Conference for Truth and Trust Online (TTO) `https://truthandtrustonline.com/`.

We also request funding at each academic site of the project, to support regular meetings of the investigating team at the universities where they are based. Regular face to face meetings are important to ensure integration of the complementary skills of the team members.

The project requests a total of £56,200 for travel to conferences and to support UK meetings between the four university sites and project partners. These funds are distributed between the four institutions thus:

- The team at Oxford (PI and PDRA 1) request funding of £11,250 for attendance at c. 5 international conferences and funding of £7,750 to attend c. 5 European conferences. Funding of £7,600 is requested for UK travel. Total: £26,600.

- The team at Lancaster (CoPI and PDRA 2) request funding of £4,800 for international travel and £3,200 for UK travel. Total £8,000.
- Birmingham (CoI) requests £3,200 for UK travel; £3,100 for European travel; £4,500 for international travel. Total £10,800
- Glasgow (CoI) requests £3,200 for UK travel; £3,100 for European travel; £4,500 for international travel. Total £10,800

Lab experiments

The project include costs of carrying out the lab experiments at Lancaster (Work Package 5). A substantial part of the proposed project involves laboratory experiments. The Economics Department at Lancaster University has state-of-the-art experimental laboratories where our experiments will be conducted, and we are applying for the marginal costs that are connected with the experiments proposed in this project. A programmer will be hired to construct the software for the experiments proposed (using Z Tree) and participants will receive performance dependent payments according to the standard practices of laboratory experiments in Economics. Our budget here involves £15,000 for lab experiments (experimental fees for around four experiments of 150 subjects each). These fees are calculated on an average payment of £25 for a 90 minute session which is considered a standard payment in UK economics laboratories. £10,000 is for programming costs, done by Computer Science students in Lancaster on an hourly basis.

Public engagement and PER training for PDRAs

As outlined in Pathways to Impact and in view of the strong public interest in the topic of fake news — and indeed other use cases of decentralised arbitration, the investigators propose to develop a public podcast or webinar, and £1,000 are budgeted for that purpose. In addition £3,000 is requested for Public Engagement with Research (PER) training for the two PDRAs.

Consumable Purchases

The two PDRAs will each need a laptop computer, providing the flexibility to work while travelling to conferences and while meeting with collaborators, and for PDRA 2 in support of the experiments. Oxford and Lancaster do not provide computer hardware for researchers, and it is normal institutional practice for research projects to fund their own computing requirements. A total of £3,000 is requested.

Dear EPSRC Panel Members,

I am writing this letter in relation to our partnership with the research team led by Professor Paul Goldberg (Oxford) and Professor Eyal Winter (Lancaster), and their research proposal Decentralised Arbitration, which they have submitted to EPSRC.

Kleros is developing an online platform to facilitate dispute resolution using online arbitrators. Our platform is based on an open code and no patented technologies.

Our interest is to form a collaboration with leading academic researchers who work in the field. The underlying ESPRC proposal which plans to, theoretically and empirically, study incentive mechanisms for decentralized arbitration falls precisely within our R&D objectives.

We have already conducted long discussions with Paul Goldberg and Eyal Winter on the topics of common interest, and on the mode of collaboration, and we are very enthusiastic about the prospects of our joint collaboration. We note that all the outcomes of our joint collaboration will be made available to the entire academic and commercial community. We hope that improving the reliability and accessibility of decentralized arbitration systems will make their use more extensive, will increase public welfare as well as the revenue for the entire industry.

Our collaboration will be two sided. We hope to affect the investigators' research focus to deal with some concrete and critical challenges faced by the current technology of decentralized arbitration. Our contributions will include:

- Access to our data

- Access to our internal research

- Access to our internal game theory researcher

- Giving a joint workshop with the project in the UK

Furthermore, we will be glad to allow the researchers to use our system and clients to design and run field experiments to test various incentives schemes for decentralized arbitration.

These experiments can substitute for some of the laboratory experiments they would otherwise perform, reduce the overall cost of the research project, and will allow the results and insights gained by the empirical part of the project to have a greater external validity.

Should the EPSRC panel have further inquiries regarding our planned collaboration within the submitted proposal for funding we will be glad to answer any question.

Sincerely,

Federico Ast, CEO, Kleros

9th of July, 2019

HM Courts &
Tribunals Service

102 Petty France
London
SW1H 9AJ

civil_and_familybusinesssupport
@justice.gov.uk

www.gov.uk

15 August 2019

Dear EPSRC Panel Members

Subject: Decentralized Arbitration Research

We are writing to you to express our support for the research proposal, Decentralised Arbitration, led by Professor Paul Goldberg (Oxford) and Professor Eyal Winter (Lancaster), which they have submitted to EPSRC.

HM Courts and Tribunals Service is working on an ambitious programme of court reform, which aims to bring new technology and modern ways of working to the way justice is administered. In collaboration with the senior judiciary and the Ministry of Justice, we are investing £1bn to reform our courts and tribunals system.

We are building a modern system for administering justice which will benefit everyone who uses it. By designing systems around the public who need and use our services, we can create a more effective system for them and generate efficiencies for the taxpayer.

In the civil jurisdiction, existing processes can be protracted, inefficient and costly. Our Online Civil Money Claims project is a digital service for people to resolve civil money claims in a simple, accessible and proportionate way, currently for claims up to £10,000 in value. This will involve more mediation and fewer hearings. It will involve simpler processes and online routes for those seeking civil justice and provide good quality digital systems to replace inefficient, paper-based systems.

As we design and develop new services, and technology, we are constantly looking for new research which can inform our work and to share insights with other organisations and projects. We have discussed both the proposed research project with the research team, and shared related research areas of interest to us, and will continue that collaboration and dialogue.

We are very interested in further research into the factors and motivations which influence an individual's decision to choose arbitration and settlement rather than formal litigation. Both to inform the information we provide individuals prior to a case being brought to court, and the continuous improvement of the mediation service we offer parties once they have brought a claim. As those services are improved and digitised we will have more opportunity, and data, for further research and collaboration.

Yours sincerely,

Kerry Greenidge (Service Manager | Civil Money Claims Project)

James Palmer (Deputy Head of Insight & User Research)

HMCTS staff are unable to provide legal advice on how you should proceed with your case. You can get advice from Citizens Advice, a local advice agency or a solicitor.

September 10, 2019

TO: Engineering and Physical Sciences Research Council
RE: Letter of support for “Decentralized Arbitration” research proposal

Dear EPSRC proposal reviewers,

I would like to offer my strongest support for the "Decentralised Arbitration" research proposal by Paul Goldberg (Oxford University), Eyal Winter (Lancaster University), and other co-investigators. The proposal is to investigate and develop a theory informing the design of systems that aim to crowdsource judgements from a network of arbitrators. The proposal mentions a number of use cases, including recognition and identification of Fake News (FN).

This topic is clearly of interest to Facebook, and myself in particular: For the last two years, I have managed the Algorithmic Game Theory and Market Design Research group in Menlo Park, where community integrity was a key areas of focus. Earlier this year I spent some time shadowing the FN detection team based in Dublin, in order to better understand how to get the right incentives in place for people whose task is to classify and analyze content shared by Facebook users. We hope the Decentralised Arbitration project will someday provide new ideas for the design of these systems.

Recently, I have transferred to Facebook’s London office, which I expect will lower frictions for any discussions with this proposal’s co-investigators. My colleague Guillaume Bouchard (also based at Facebook's London office) and I have already had very productive meetings with Prof. Goldberg discussing the technical issues and related ideas. In these early stages of the research project, I am looking forward to providing sanity checks about the underlying assumptions being made, to help increase the likelihood that the research may someday be applied in practice. As the project progresses, I am happy to connect the co-investigators with relevant product teams at Facebook, so that they can explore the possibility of a closer collaboration with those teams as appropriate.

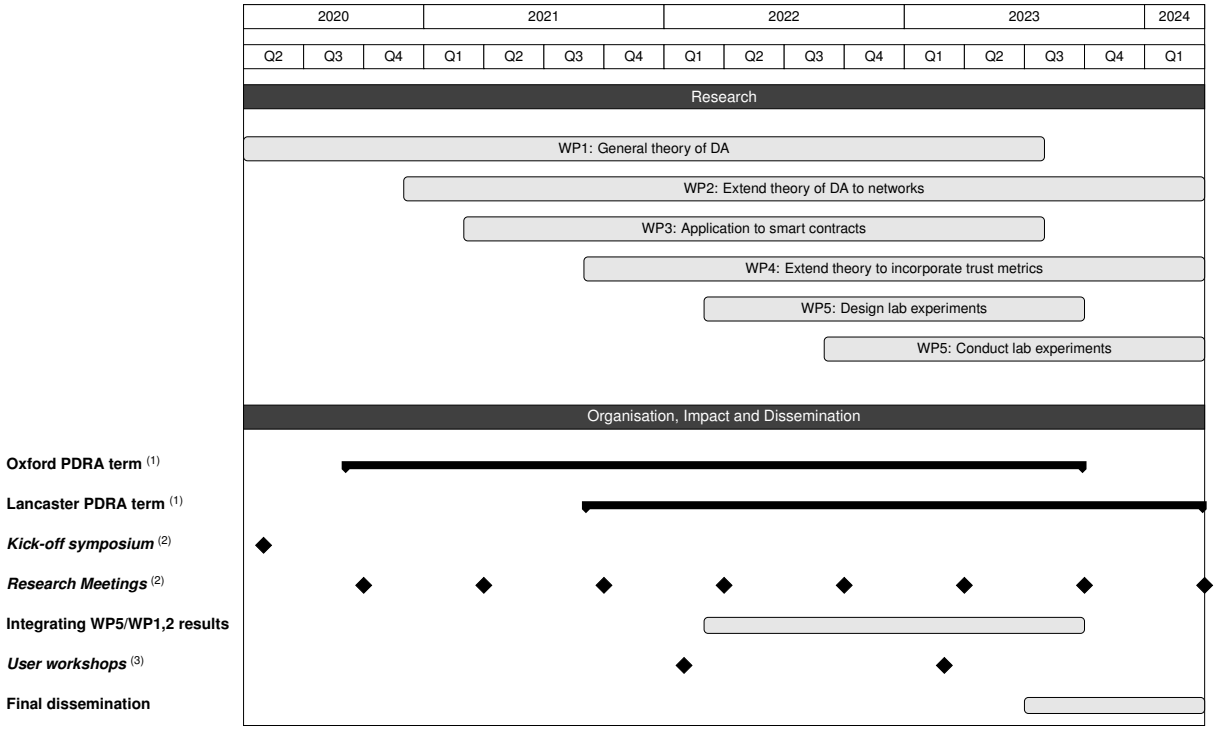
If you have any questions, please don’t hesitate to contact me.

Sincerely,

Eric Sodomka
sodomka@fb.com
+44 07778 299621

Address: 1 Hacker Way
Menlo Park, CA 94025

Workplan: Decentralised Arbitration



Notes

(1) PDRAs at Oxford and Lancaster will typically have weekly meetings with Prof. Goldberg and Prof. Winter respectively. We have requested funds for the PDRAs to regularly travel to the other sites for further collaboration.

(2) All project partners will participate in the kick-off symposium and regular meetings.

(3) Users include the writers of the support letters in addition to other outside users.

Members of the team will travel to conferences and workshops for the purpose of dissemination of results; the exact timing of these meetings is not fixed.

Risk mitigation For the theory-oriented Work Packages 1–4, the general risk is inability to prove satisfying performance guarantees of DA systems of interest. The general approach to deal with this risk is to identify more demanding assumptions to be made on participants on a DA network, that would allow valuable conclusions to be drawn. It then becomes important to check whether any such assumptions are realistic: our project partners are of considerable value here, for the purpose of checking that use cases are being modelled in an acceptable way.

There is a risk that experimental data from WP5 may disagree with modelling assumptions. Experiments in WP5 will be informed by the analysis of WPs 1,2; we will be able to design further experiments that are informed by the results of previous ones.

With regard to getting from theory to practical applications, flexibility may be needed. The status of truthful (Vickrey-Clark-Groves) auctions serves both as warning and inspiration. Despite the rich theory of truthful auctions, it is unusual to see them in practice [17,27]. On the other hand, the closely related *generalized second-price (GSP) auction* exploits the ideas of truthful auctions and is used in Google's adwords technology.