Macartan Humphreys, Berlin

Alexandra Scacco, Berlin

Julio Saul Solís Arce, Berlin

Rolling Metaketa: A strategy for theory-based cross-replication

Project Description

1 Starting Point

1.1 State of the art and preliminary work

In recent years, social scientists have grown increasingly concerned about a set of challenges that make it difficult to learn across studies—including heterogeneity in study design, effect heterogeneity, and selective reporting—and ongoing challenges for establishing the external validity of empirical findings (Dunning et al. 2019, 16-25). These are features of the so-called "replication crisis."

One approach designed to address many of these challenges is the "Metaketa" model for coordinated trials, pioneered by the Evidence in Governance and Politics (EGAP) network. This approach, which so far has mostly been applied in political science, involves a set of coordinated field experiments that are carried out in multiple sites. The experiments are designed in advance to address a common research question, with a commitment to a common design (for e.g., a common treatment arm) and analysis standard. After the individual study interventions are completed, harmonized measurement of inputs, and outcomes across studies facilitates meta-analysis. Ultimately, such meta-analyses advance cumulative learning, enabling social scientists to gather knowledge from multiple sites on a single topic, while maintaining high standards in research transparency and reproducibility, through open data and code, third-party data analysis and pre-registration of designs and analysis plans. (Dunning et al. 2019, 25-30). The Metaketas inherently involve two forms of replication: *analytic replication* of each study and simultaneous *field replication* across studies—or what might be called "cross replication." To date there have been five such Metaketas fielded (https://egap.org/our-work/the-Metaketa-initiative/).

Alongside its many strengths, there are two limitations of the Metaketa model as currently practiced that this project seeks to address (Dunning et al. 2019, 31).

The first comes from the degree of coordination required under the current model. Under the standard model, individual projects are jointly financed and implemented in parallel. Single-source financing can result in funders having a larger influence on topic choice and potentially excludes a wide number of subjects relevant to the scientific community, which are not necessarily aligned with policy priorities or funders' interests. For the same reason, Metaketas do not connect naturally with "normal" science processes. For instance, most scientific efforts are not centrally coordinated; they have lengthy and non-parallel timeframes. Currently, knowledge generated by such processes cannot easily feed into the knowledge accumulation so central to the Metaketa ambitions.

The second is the backseat role played by theory in existing Metaketas. A great advantage of experimental designs is that they make it possible to estimate treatment effects from samples without recourse to elaborate theories. This makes it possible, in principle, to coordinate research without a common theoretical framework. However, "transportation" of findings to new contexts or different treatments does require theory in order to take account of relevant differences between study populations and target populations (Bareinboim and Pearl 2013,

Wilke and Humphreys, 2020). Moreover, social scientists often engage in micro-level research (such as the Metaketa experiments) in order to learn about macro-level questions. Yet, drawing inferences between levels in this way is generally not possible without reference to a theoretical structure (Humphreys and Scacco 2020).

These two issues are linked in an important way: embedding related research in a theoretical structure that can take account of—and facilitate learning about—heterogeneity, can in principle allow for the incorporation of knowledge produced by less coordinated, asynchronous, research processes.

The goal of our project is to address these frontier issues and extend the concept of EGAP's Metaketa by working to: (i) set up a structure for such a theory based "Rolling Metaketa" to aggregate findings as they are produced in real time and (ii) show proof of concept by implementing the approach for an important research question that has generated a large body of experimental studies across the social sciences: *how does intergroup contact affect prejudice*?

Current approaches for bridging theory and experiments

Critics have pointed to the disconnect between theory and experimental evidence (Card, DellaVigna, and Malmendier 2011; Deaton and Cartwright 2018; Harrison 2014; Huber 2017). This disconnection is relevant, because researchers work with theories, even if only in the back of their minds, to motivate the design of an experiment. Moreover, bringing together theory and experiments is fundamental to the scientific process, since experimental evidence could be used to select, revise or discard theories.

Recent initiatives have pushed for greater theory specification in applied research. Glöckner, Fiedler, and Renkewitz (2018) suggest the creation of a "theory database", where authors can outline their theory in detail and store it in an open repository, so that theories are opened for scrutiny. Along the same lines, Gray (2017) and others have started to implement these ideas by specifying theories in the form of "theory maps", a form of diagram that links theoretical constructs to each other. The DeclareDesign project (Blair et al. 2019) provides a framework for including a formal specification of a theory, in code, as part of research design registration. Advances in data fusion approaches (Hünermund and Bareinboim, 2019) give a powerful framework for using causal models to justify extrapolation of inferences, which presupposes theory articulation.

These are very promising approaches and bring transparency and prominence to the role of theory. In large part, however, they are focused on improving individual studies, and how rival theories fare against each other. The approach we propose shares with Glöckner, Fiedler, and Renkewitz (2018) a commitment to theoretical learning *across* studies but differs to the extent that, rather than pitting theories against each other, the focus is on learning about overarching theories.

A second welcome development has been a growth in interest in structural models as a way to better integrate theories into analysis. These include economic structural models and non-parametric structural causal models. Both approaches can help articulate mid-range theories that can guide applied research (Wilke and Humphreys, 2020). Organizations like the UK's Centre of Excellence for Development Impact and Learning (CEDIL) are currently promoting similar approaches to integrate "middle-level" or "middle-range" theories in their studies in order to later implement process-tracing strategies that yield evidence of the relevant mechanisms driving the effects of certain policy interventions (Davey et al. 2018).

Many of these approaches focus on how theory can aid empirical analysis or be tested by

empirical analysis but focus less on how empirical analysis can be used to update theories. Bayesian structural causal models in particular, we believe, provide, a useful framework to both specify a theory and use data to update the theory. These models are rich enough to incorporate the estimands that are the usual focus of design based inferential strategies—such as average treatment effects—and include contextual and process variables in a natural way—such as mediators and moderators. Moreover, they can themselves be derived from underlying behavioral models (see, for instance, Low and Meghir (2008) on structural economic models).

A conceptual advantage of focusing on structural causal models is that the focus of research shifts from understanding *effects* in a given place to developing an *explanation* that—if valid—has implications across contexts. Causal models present great advantages for model-based meta-analysis. We know, for instance, that in a fully specified structural causal model, every counterfactual quantity is computable. By rules of the causal models, we can predict the effects of interventions, derive forms of mediation analysis, and "transport" results to target domain with the help of causal knowledge in a source domain (Pearl and Bareinboim 2011). All of these features of causal models mean that there is the possibility to combine causal knowledge from several different source domains (Bareinboim and Pearl 2013) and even combining experimental and observational data (Bareinboim and Pearl 2014). While some approaches are being developed on the lines of transportability of results, the use of causal models for the purpose of meta-analysis remains underdeveloped (Dahabreh et al. 2020).

State of the art on coordinated trials, meta-analysis, and cross replication.

We also build on recent developments in meta-analytic methods.

Data driven strategies. The growth of machine learning techniques, including BART and random forests procedures, opens the possibility of addressing heterogeneity by systematic searches, and cross validation, of conditions under which causal relations are more or less likely to hold. Network meta-analyses have also been used to compare heterogeneous populations subject to different combinations of treatments in an experiment. They can be flexible enough to get indirect estimates of treatments contrast even when no RCT has directly compared two treatments directly (Dias et al. 2018), which is a counterfactual estimation of sorts. New efforts seek to connect these approaches to approaches targeting causal inference (e.g., Schnitzer 2016), with striking new work using machine learning to try to assess mediation processes (Farbmacher et al 2020). We see the goals of these data driven approaches as largely parallel to the goals of our project and propose incorporating these as rival methods to help assess the merits of our approach in Component III of our workplan (see below).

Model based meta-analysis. There is an older and a new literature emerging on more model centered approaches to meta-analysis that involve structural equation modelling (Becker 2009). A drawback is that in many applications model forms are often restricted, for instance with linearity imposed and a reduced scope to learn abbot interactions and counterfactuals. The approach to meta-analysis via hierarchical models described in Gelman et al. (2013) also represents a model-based approach to the extent that it targets superpopulation parameters. An important insight from these approaches is that in such a meta-analysis, as one learns about superpopulation parameters one also reevaluates study level estimates, separating fundamental variability from sampling error. This insight is one that we incorporate also in the hierarchical causal models we employ, with a difference that the parameters of interest are not average effects but rather the structure causal relations.

Meta-analysis by Design. Traditionally, meta-analytic strategies take the body of evidence as given and seek to aggregate knowledge from it. Systematic review procedures put in place admissibility conditions and seek to ensure representativeness of bodies of research, but the

population of interest is generally existing research rather than possible domains of application. In some cases, studies in a meta-analysis have a natural compatibility, emerging from common research agendas, even if not designed specifically for meta-analysis (Banerjee et al. 2015). Yet there has been a growing interest in strategies, like the Metaketas, to implement projects in a way that facilitates meta-analysis at future stages. Maniadis et al. (2014) make a case for replications of this form in economics. Henrich et al. (2004) is an influential early study that goes further in systematically assessing variation in findings across cases that were coordinated in advance. The "ManyLabs" projects in psychology are a striking example of large-scale coordination with both sequential and parallel replication (Open Science Collaboration 2015). To our knowledge, however, meta-analyses to date that use studies designed for purpose have not been designed specifically to assess heterogeneity using a common causal model that they jointly seek to update. Our project which combines both analysis of existing work and meta-analysis by design will seek to combine insights from these two approaches.

Preliminary work on coordinated trials and inference integration.

At the Institutions and Political Inequality Unit at WZB Berlin Social Science Center, we have engaged in considerable preliminary work both in terms of setting up structures for aggregating research findings and in canvassing research in the substantive field of intergroup contact.

Coordinated trials: We contributed to the "Metaketa" initiative housed at the Evidence in Governance and Politics (EGAP) network. The Metaketa initiative was set up while one of our PIs (Humphreys) was executive director of EGAP. Humphreys was also a PI on Metaketa I and a member of the Metaketa steering committee. Researchers at WZB also contributed to the meta-analysis of Metaketa I and produced a user ("shiny") interface that lets users replicate in real time, and explore analyses and robustness of findings to the inclusion or exclusion of different studies. See https://egap.shinyapps.io/Metaketa_shiny/. Similar interactive tools will be used for the Rolling Metaketa.

Coordinated field measurement. We are leading the meta-analysis for a multi-country collection of studies coordinated by African Programming and Research Initiative to End Slavery (APRIES) to assess the prevalence of human trafficking prevalence estimation. In this study we pioneer simultaneous cross-replication approaches for complex multisite measurement strategies. The strategy, in development, is available at https://macartan.github.io/hidden_populations/. We have also been working with multiple teams to gather harmonized data on emerging impacts and responses to the Covid pandemic. A tool we have developed for coordinating measurement and analysis of attitudes to Covid vaccines can be seen at https://wzb-ipi.github.io/covid_vaccines/.

Integrated inferences: We developed the CausalQueries R package that lets users define and aggregate causal models. The package is publicly available via the Comprehensive R Archive Network (CRAN). We prepared a guide for examples with applications to combine inferences from qualitative and quantitative analyses, inferences from observational and experimental studies, and inferences from multiple trials examining different parts of a common causal model. See https://macartan.github.io/causalmodels/. CausalQueries shows that the inference aggregation methods that will compose the Rolling Metaketa are feasible. Extensions to CausalQueries now include functionality for hierarchical causal models; which represents a combination of approaches often seen as rival.

Meta-models: In addition to studies that apply common models to heterogeneous populations researchers in the group have been developing the tools to aggregate insights from rival models applied to the same data. In particular, in November 2020 we launched a challenge to crowdsource predictive models for cross-national Corona mortality. Predictions from submitted models are to be combined using a stacking approach–an approach to combining models that

are all operating on the same data. For more see: https://eos.wzb.eu/ipi/shiny-covid-crowdsourcing/

These projects demonstrate experience with the high level of research collaboration needed for a project like the Rolling Metaketa and display the type of innovation that will be needed to develop and test each of the above strategies.

Preliminary work on assessing social contact interventions

For our first application of the Rolling Metaketa, we will focus on studies of intergroup contact – a field that has generated hundreds of studies across the social sciences since Gordon Allport first outlined the "contact hypothesis"—that equal status contact in pursuit of common goals should reduce prejudice across even deep social divides—in *The Nature of Prejudice* in 1954. We have undertaken extensive preliminary work both in designing and implementing field experiments testing the impact of intergroup contact in conflict settings, and in canvassing a broad set of existing studies.

Although there is a large empirical literature examining the contact hypothesis (Pettigrew and Tropp 2006), until recently, the number of studies employing field experimental approaches in settings with marked intergroup prejudice and discrimination has been surprisingly limited (Paluck, Green, and Green 2019). A recent exception is work by Alexandra Scacco and Shana Warren (Scacco and Warren 2018) who designed and implemented an innovative education-based field experiment—the Urban Youth Vocational Training (UYVT) program—to assess how sustained contact between members of groups in conflict can reduce prejudice and discrimination, increase cooperation, and limit support for violence. The program was implemented in the Kaduna in northern Nigeria, as setting that had experienced deadly Christian-Muslim clashes over the past decade.

Contrary to expectations from the contact literature, Scacco and Warren find no meaningful changes in prejudice after four months of Christian-Muslim contact in a positive educational setting, but find a significant and robust reduction in discriminatory behavior among both Christian and Muslim participants assigned to the social contact treatments (religiously mixed classrooms). This finding can be seen as a failure to replicate results of previous studies but also, through the discovery of a new pattern, a replication challenge for future studies.

In a second ongoing study, Scacco, together with Salma Mousa (Stanford, Immigration Policy Lab), is currently in the process of implementing a field experiment in Lebanon that brings together hundreds of Syrian refugees and Lebanese youths for several months of soccer practice and league play in order to gauge the effect of intergroup contact on prejudiced attitudes, discriminatory behavior and norms about appropriate intergroup interactions. The study randomized the composition of soccer teams (Syrian-Lebanese, or homogeneous Lebanese and Syrian teams), and whether supplementary intervention programming content focuses directly on intergroup cooperation or health and nutrition-related issues. In addition to baseline and endline survey data, qualitative information is also being collected to carefully trace potential causal mechanisms.

Scacco is currently in the process of fielding several other field experiments related to intergroup contact in settings of protracted social conflict. Two examples illustrate the type of work in progress. First, a field experimental study with Rebecca Littman (University of Illinois, Chicago) and a team of social psychologists investigates whether "virtual contact", in the form of a positive relationships between Christians and Muslims depicted in a popular Nigerian television series, can improve interreligious attitudes and behavior. A second experimental study currently in the field investigates whether exposure on Facebook to personal-interest news about (the often difficult) circumstances of daily life experienced by Arab residents of East

Jerusalem shifts prejudiced attitudes and online behaviors of Jews living in West Jerusalem.

In developing these projects, we have been working with an expanding network of scholars at the cutting edge of empirical research on intergroup relations. In particular, Scacco has founded, together with Rebecca Littman, Chagai Weiss (University of Wisconsin-Madison), and Cara Wong (University of Illinois at Urbana-Champaign), the Intergroup Relations Working Group (IGR)—an open interdisciplinary network engaging psychologists, economics, political scientists, and sociologists in discussions of new work in this area. More information on this network is available from its website https://igr.netlify.app/, including a list of recent presentations of new experimental on intergroup relations.

As a final preparatory step, we hosted a conference in 2018 at the WZB that brought together a diverse group of scholars and practitioners working on social contact to outline common core elements of research designs and identify open questions in the field. The ideas from this workshop were presented at the EGAP meetings in Wageningen, the Netherlands, and inform the current proposal.

This preliminary work establishes, in our view, that there is an emerging body of high-quality experimental results in this area and a large community of scholars with a common interest in seeing their findings replicated, reconciling conflicting results, and contributing to meta-analytic research that can help explain heterogeneity across studies.

1.2 Project-related publications

1.2.1 Articles published by outlets with scientific quality assurance, book publications, and works accepted for publication but not yet published.

Related to replication, aggregation, transparency:

1. Blair, G, J Cooper, A Coppock, and M **Humphreys**. "Declaring and diagnosing research designs." *American Political Science Review* 113, no. 3 (2019): 838-859.
2. Dunning, T., Grossman, G., **Humphreys**, M., Hyde, S. D., McIntosh, C., & Nellis, G. (Eds.). (2019). Information, Accountability, and Cumulative Learning: Lessons from Metaketa I. Cambridge University Press.
3. **Humphreys**, M., De la Sierra, R. S., & Van der Windt, P. (2013). Fishing, commitment, and communication: A proposal for comprehensive nonbinding research registration. *Political Analysis*, 1-20.
4. **Humphreys**, M., & **Scacco**, A. (2020). The Aggregation Challenge. *World Development*, 127, 104806.
5. **Humphreys**, M. & **Wilke**, A. (2020). "Field Experiments, Theory and External Validity" in *Handbook of Research Methods in Political Science and International Relations*. R Franzese and L Curini (Eds) Oxford UP.
6. B. A. Nosek, G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, C. D. Chambers, G. Chin, G. Christensen, M. Contestabile, A. Dafoe, E. Eich, J. Freese, R. Glennerster, D. Goroff, D. P. Green, B. Hesse, M. **Humphreys**, J. Ishiyama, D. Karlan, A. Kraut, A. Lupia, P. Mabry, T. Madon, N. Malhotra, E. Mayo-Wilson, M. McNutt, E. Miguel, E. Levy Paluck, U. Simonsohn, C. Soderberg, B. A. Spellman, J. Turitto, G. VandenBos, S. Vazire, E. J. Wagenmakers, R. Wilson, T. Yarkoni. "Promoting an open research culture" *Science* 2015: Vol. 348, Issue 6242, pp. 1422-1425

Related to social contact:

7. Habyarimana, J., **Humphreys, M**., Posner, D. N., & Weinstein, J. M. (2007). Why does ethnic diversity undermine public goods provision? *American Political Science Review*, 709-725.
8. Scacco, A., & Warren, S. S. (2018). Can social contact reduce prejudice and discrimination? Evidence from a field experiment in Nigeria. *The American Political Science Review,* 112(3), 654-677.

8.2.2 Other publications, both peer-reviewed and non-peer-reviewed

9. Bicalho, C., Cooper, J., **Humphreys**, M., Jacobs, A., Heidemanns, M., Medina, L, **Solis**, J., Syunyaev, G. (2020). CausalQueries: Make, Update, and Query Binary Causal Models. R Package. https://cran.r-project.org/web/packages/CausalQueries/index.html

10. Davey, Calum, James Hargreaves, Syreen Hassan, Nancy Cartwright, Macartan **Humphreys**, Edoardo Masset, and C. Bonell. "Designing evaluations to provide evidence to inform action in new settings." CEDIL Inception Paper 2 (2018). https://cedilprogramme.org/wp-content/uploads/2018/10/Designing-evaluations-to-provide-evidence.pdf

2. Objectives and work programme

2.1 Anticipated total duration of the project

Following the call for proposal of this SPP Program, we intend to run our project for 3 years (36 months). See details on timing of the project below.

2.2 Objectives

There are three major components to the work, each with specific objectives.

Component I: Developing methods for aggregating research findings from studies in real time using an encompassing causal model:

a. Developing tools for model-based meta-analysis, including testing approaches for generating common causal models.
b. Setting up an infrastructure that guides the design of future studies to maximize our ability to incorporate new results in a rolling meta-analysis.
c. Developing the interactive meta-analysis tools in a way that they could be easily adapted for different group wanting to launch their own Rolling Metaketa.

Component II: Field testing the approach with an application to studies that examine variants of the "contact hypothesis" (Allport 1954 [1979], 251). As noted above, the contact hypothesis has inspired the work of social psychologists for decades and has recently been take up by political scientists and economists. We believe this is a good first application to field test the approach. There exists a recent body of largely comparable studies, heterogeneity of contexts of study, and heterogeneity of findings. As described by Paluck, Green, and Green (2019) however, there is a fundamental "lack of research that systematically investigates the scope conditions suggested by Allport (1954) under which contact is most influential."

By specifying the contact hypothesis as a causal model and implementing a Rolling Metaketa around it we aim to:

a. Replicating existing work and set up a structure for ongoing replication of future work in the area.
b. Updating a covering model that allows us to learn about the sources of heterogeneity and assess the extent to which failures to cross-replicate can be explained by a covering theory.
c. Provide proof of concept for a structure for open ended theory based meta-analysis and cross-replication.

Done right, this would have major consequences for research and development interventions promoting intergroup contact. But it would also stress test an approach to meta-analysis that is theory-based and flexible enough to incorporate emerging work.

Component III: Critiquing, refining, and disseminating results. Our goal here is to clarify strengths and weaknesses of the approach and share the lessons learned with different researchers and institutions around the globe both through presentations but also by generating an open source repository of tools. A possible result of this review stage is that, satisfying as the explanations from a covering structural causal model are, the inferences from the model are no stronger than those from a more data-driven approach.

How are the objectives of this project related to the META-REP SPP?

In terms of the questions posed on by META-REP, we believe our project is especially suitable to address the "Why" question (i.e., Why Do Replications Fail, and Why Do Replication Rates Vary Within and Across Disciplines?) and does it in key ways that adhere to the specification of the SPP call.

The call identifies two candidates for this why question: (i) inflated false positive and (ii) underappreciated context dependency of effects. The Metaketa approach seeks to control (i) by design and in doing so be in a strong position to assess (ii).

The key strategies to control (i) are pre-registration requirements and third-party analytic replication. We will naturally be keen to learn from the broader META-REP project about the effectiveness of these and related strategies. Given (i), we assess (ii) by using a structure to assess the role of heterogeneity within the context of a common causal model. We think of these models as "covering" in the sense that they can nest study specific models.

In practice this combines three forms of replication:

1. Populating the Meta-analysis with prior studies requires replication and re-analyses of these studies.
2. Following Metaketa standards, admittance of new studies also involves third party replication, coupled with reconciliation with pre-analysis plans.
3. By employing a covering model, specific findings from different studies can be assessed using data from other studies. The structured Meta-analysis thus engages in a form of "cross replication"— where inferences in one case are updated and reassessed in light of findings in another.

The contribution to META-REP goals is to integrate these forms of replication in a way that:

- can address the why question posed by the call
- is *constructive* such that rather than being perceived as a failure, variation in results themselves shed light on heterogeneity in underlying process, and
- provides a framework to "mainstream" replication of this form so that it gets built in to research agendas around major questions.

2.3 Work programme including proposed research methods

We describe the expected timing, specific activities, staffing, and methods used in turn.

Timing

The major first phases of the project are shown in Figure 1. A goal of the study is to create a platform for ongoing meta-analysis that continues beyond the timeline of the project. This is indicated by the arrow in the Table. A part of Component III however is to assess procedures that make this possible, whether by "housing" the Metaketa with a journal, research network, or open science platform.

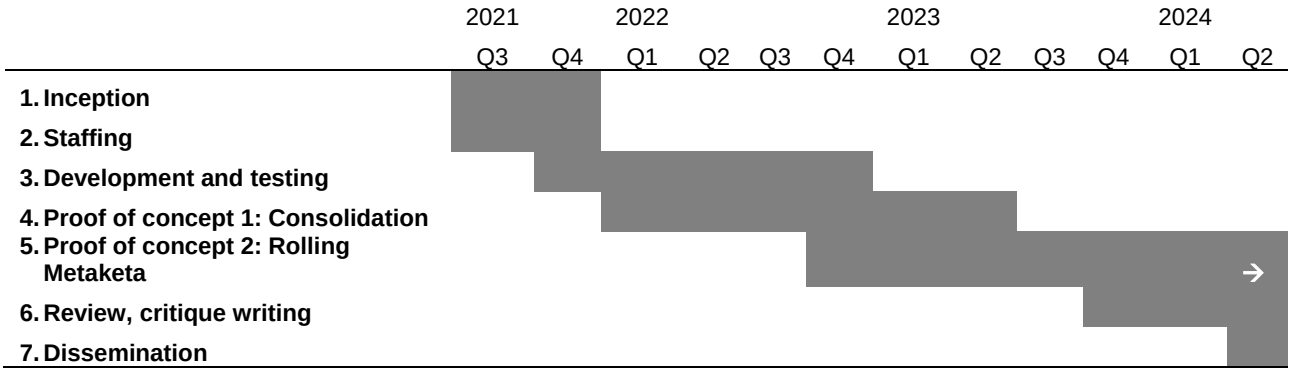

Figure 1: Timing of activities

Activities

Table 1 summarizes the core activities associated with each of these phases and relates them to project objectives.

Table 1: Components and activities

Phase	Activities	Objectives
1. Inception	• Draft and circulate a white paper setting up the principles of the Rolling Metaketa, soliciting feedback. • Develop and set up a beta website interface that shows the core functionality of the Rolling Metaketa framework. Likely done with a "mock" theory and data and assessed using a hierarchical causal model building on the package CausalQueries.	**I Development**
2. Staffing	• Open cross disciplinary call to hire a postdoctoral fellow. • Set up the scientific committee. • Commence preparations for the organization of a workshop with the scientific committee to discuss the strategy and design.	**I Development**
3. Development and testing	• Conduct a literature review on field experiments of intergroup contact. • Gather data from original studies found in the literature review. • Replicate meta-analysis from Pettigrew and Tropp (2006) and Paluck, Green, and Green (2019). • Draft procedural guidelines with admission criteria to the Rolling Metaketa (see below). • Design workshop with scientific committee.	**I Development**
4. Proof of concept 1: Consolidation	• Develop covering structural causal model. • Map from model to nodes to data available in individual studies. • Determine admissible set of studies. • Replicate individual studies. • Update covering model.	**II Field Testing**
5. Proof of concept 2: Rolling Metaketa	• Canvass scholars preparing and conducting new studies. Develop material to facilitate harmonization of designs and measures. • Integration of new studies as they become available. • Update and refine covering model.	**II Field Testing**
6. Review,	• Model checking analyses: implementing tests of the	**III Review and**

critique writing	covering model and performance relative to rival strategies via cross validation. • Paper 1: Draft and circulate paper 1 providing the results of the Social contact meta-analysis in academic seminars. Goal here is to work with a journal willing to innovate in publishing a "live" live where the meta-analysis of record updates in real time. • Paper 2: Draft and circulate paper 2 reporting on strategies of inference aggregation and lessons from the Rolling Metaketa. • Set up a skeleton of the project repository for use by future Rolling Metaketas.	**Dissemination**
7. Dissemination and wrap up	• EGAP workshop hosted at WZB to report lessons and identify future applications of the approach. • Support efforts by other groups to set up Rolling Metaketa structures. • Explore procedures for ongoing review of new entries to the Rolling Metaketa that will take place after project completion.	**III Review and Dissemination**

Staffing and responsibilities

Key to implementation is that hiring of a postdoctoral fellow who will engage on all components of the project in charge and take primary responsibility for Component II. The position will be held for the entire duration of the program (36 months). Additionally, we will set up a scientific committee to advise on the methods developed for the causal meta-analysis and the implementation of the social contact Rolling Metaketa. Below we make explicit the roles of each of these persons in executing the project.

PI roles: All PIs will collaborate on all components of the project. Humphreys and Scacco will jointly coordinate the project with Humphreys and Solis-Arce taking primary responsibility for method development and infrastructure in Component I and Scacco and the post-doc hire taking primary responsibility for designing and implementing the proof of concept in Component II.

The **postdoctoral research fellow** will be recruited from academics holding a doctoral degree (or in process of presenting their dissertation) with interests in meta-analysis and causal models, but with substantive expertise in contact between social groups and/or strategies of prejudice reduction. The responsibilities of the postdoctoral fellow are to:

- Lead, together with PI Scacco, the implementation of the first Rolling Metaketa: Intergroup contact.
- Research and develop the necessary methods of inference aggregation used in the causal meta-analysis and the transportability section of the Rolling Metaketa.
- Conduct the literature review previous to a meta-analysis.
- Communicate with scholars and retrieve data from primary studies to set the meta-analysis.
- Coordinate with technical staff in setting up the user interface of the Rolling Metaketa.
- Consult and collaborate with the scientific committee to refine the inference aggregation methods.
- Contribute to analysis and drafting of research products.

The **scientific committee** will be in charge of overviewing the methods developed for the Rolling Metaketa as well as providing substantive feedback to the social contact application. They will provide advise towards the scientific foundations of the framework and guidance in

terms of revisions and enhancements. They will include:

- experts in inference aggregation, including leads in the EGAP Metaketa initiative, and experts in hierarchical modelling and structural modelling.
- experts in social contact from multiple disciplines (psychology, economics, political science, statistics, computer science), including researchers active in the Inter Group Relations network
- practitioners from agencies that employ academic research to guide design of implementations.

The underlying idea is that method and substantive-minded people come together in the space to develop a common working framework. We will commission a small set of texts from a subset of them for a launching workshop. Further, we make available a small “summer grant” to receive key members of the scientific committee in Berlin to provide expert advise in the development of the project.

We have already secured agreement from experts in all three of these groups (see section 6.4 and 6.5). If possible, we hope to include key members from other META-REP projects among the first group of experts.

Methods and tools

Multiple methods and tools are used at different points in the project. We first outline the likely user facing interface for the Rolling Metaketa which, will likely make use of R shiny tools and GitHub and OSF project management tools. We then describe the methods that are used to power the interface and later in our own analysis.

Interface

We describe the generation of a template for admitting and aggregating findings during the development phase. Still to be developed, we imagine an interface with multiple "panels" including:

- ***1) Theory specification***: in which a theory of relevance to the social and behavioural sciences is shown as directed cyclical graph from a causal model with accompanying annotations from relevant literatures.

- ***2) Meta-analysis:*** in which commonly used methods of meta-analysis replicate individual studies and implements meta-analysis. This panel should be powered by an interactive tool (e.g., R Shiny) that allows for different types of meta-analysis (fixed effects, random effects) and specifications to run the meta-analysis (study-level moderators, sub-samples, etc.). The interface we developed for Metaketa I provides an example of how this panel might look: https://egap.shinyapps.io/Metaketa_shiny/

- ***3) Causal meta-analysis*****:** in which methods of causal models are used to pass on data from experiments into the DAG. Using causal models allow us to consider a situation in which we believe the same theory holds in multiple sites but in which learning about the model requires combining data about different parts of the model from multiple studies, with studies potentially looking only at specific parts of the theory. The interface should provide tools to allow users to ask "causal queries" of the model. Other interactive features should permit the customization of data passed to the model. For instance, users could select a subsample of studies to be passed on to the model. Potentially, users could also provide alternative specifications of the theory and pass on the same data. This could allow for theory revision or "causal discovery" (try to learn compatible model from conditional independence relations found in the data).

- ***4) Transportability***: in this tab, we provide the opportunity to users to use the aggregated data of studies compiled to understand the relative contribution of a potential new study. Using a framework like that developed by Dahabreh et al. (2020) we permit users to pass on their own baseline data or simulate a target population to derive expected causal effects of an intervention similar to previous studies. This could be used to enhance the design of a new study, maximize power of the running theory, or simply understand how a new study would affect the current research synthesis shown in tabs 2 and 3 (e.g., a form of sensitivity analysis).

- ***5) Challenges blog:*** in which experts in the field of the theory (experimentalists and non-experimentalists) can provide input regarding the theory and the type of experimental interventions that are being done to learn from the theory. Discussions could focus on design, relative pathways of the model, mediation analysis, theory specification, etc. They are challenges because they should orient the implementation of new experiments based on the research synthesis of the previous tabs.

- ***6) Add my study:*** instructions and inclusion criteria are given to authors to contact coordinating team in order to keep the Metaketa "rolling" by passing new data from the latest studies as they come from the field. As described above study addition will require prior establishment of admittance criteria. These will likely involve registration and open science criteria for new studies as conditions for admittance.

Methods for model construction and meta-analysis

Model development in Panel 1 will make use of methods developed by Pearl (2011) and colleagues and make use of functionality developed in the Dagitty package.

Analysis in panel 2 will use meta-analytic tools and likely involve frequentist approaches – random and fixed effects as well as Bayesian drawing particular on Beath and Bolker. (2015) and Gelman et al (2013). An instance of comparisons of these models is described by Humphreys working with the declare design team at: https://declaredesign.org/blog/2018-12-11-meta-analysis.html.

Panel 3 will likely make use of Bayesian updating of causal models as implemented by the CausalQueries package. Alongside CausalQueries, Analysis in Panel 4 would use procedures developed by Barenboim and colleagues as part of the Data Fusion project (Hünermund and Bareinboim, 2019).

Methods for model evaluation

Component III involves a critical and a promotional aspect. An obvious weakness of theory-based approaches is that, because inferences are made conditional on a theory, inferences can be invalid when theories are misleading. Fortunately, with access to data from multiple studies we have an ability to assess the extent to which claims are theory dependent, the extent to which data patterns are inconsistent with theoretical expectations, and the extent to which one theory outperforms another. In the context of the Bayesian structural causal models that we expect to employ we can use, for instance, cross validation approaches, tests of conditional independence relations (Zhang et al 2012), and checks on posterior predictive fits (Gelman et al 1996). In particular, using cross validation we can assess the extent to which out-of-sample predictions from causal models are better or worse than what can be developed from less theoretical approaches--- such as machine learning approaches for identifying relevant dimensions of heterogeneity.

3 Bibliography concerning the state of the art, the research objectives, and the work programme

Allport, Gordon W. 1979. *The Nature Of Prejudice: 25th Anniversary Edition*. Basic Books.

Banerjee, Abhijit, et al. 2015. "A multifaceted program causes lasting progress for the very poor: Evidence from six countries." *Science* 348.6236.

Bareinboim, Elias, and Judea Pearl. 2013. 'Meta-Transportability of Causal Effects: A Formal Approach'. In *Artificial Intelligence and Statistics*, 135–43. PMLR. http://proceedings.mlr.press/v31/bareinboim13a.html.

———. 2014. 'Transportability from Multiple Environments with Limited Experiments: Completeness Results'. *Advances in Neural Information Processing Systems* 27: 280–88.

Beath, Ken J. and Ben Bolker. 2015. "Metaplus: Robust Meta-Analysis and Meta-Regression." R Package Version 0.7-4.

Becker, Betsy Jane. 2009. "Model-Based Meta-Analysis." In The Handbook of Research Synthesis and Meta-Analysis", 2nd edition, ed. Harris Cooper, Larry Hedges, and Jeffrey Valentine, 377-398. New York: Russell Sage Foundation.

Blair, Graeme, Jasper Cooper, Alexander Coppock, and Macartan Humphreys. 2019. 'Declaring and Diagnosing Research Designs'. *American Political Science Review* 113 (3): 838–59. https://doi.org/10.1017/S0003055419000194.

Card, David, Stefano DellaVigna, and Ulrike Malmendier. 2011. 'The Role of Theory in Field Experiments'. *Journal of Economic Perspectives* 25 (3): 39–62. https://doi.org/10.1257/jep.25.3.39.

Dahabreh, Issa J., Lucia C. Petito, Sarah E. Robertson, Miguel A. Hernán, and Jon A. Steingrimsson. 2020. 'Toward Causally Interpretable Meta-Analysis: Transporting Inferences from Multiple Randomized Trials to a New Target Population'. *Epidemiology (Cambridge, Mass.)* 31 (3): 334–44. https://doi.org/10.1097/EDE.0000000000001177.

Davey, Calum, James Hargreaves, Syreen Hassan, Nancy Cartwright, Macartan Humphreys, Edoardo Masset, Chris Bonell, Sandy Oliver, Audrey Prost, David Gough. 2018. Designing evaluations to provide evidence to inform action in new settings. CEDIL Inception Paper, 2.

Deaton, Angus, and Nancy Cartwright. 2018. 'Understanding and Misunderstanding Randomized Controlled Trials'. *Social Science & Medicine*, Randomized Controlled Trials and Evidence-based Policy: A Multidisciplinary Dialogue, 210 (August): 2–21. https://doi.org/10.1016/j.socscimed.2017.12.005.

Dias, Sofia, A. E. Ades, Nicky J. Welton, Jeroen P. Jansen and Alexander J. Sutton. 2018. Network meta-analysis for decision-making. John Wiley & Sons.

Dunning, Thad, Guy Grossman, Macartan Humphreys, Susan D. Hyde, Craig McIntosh, and Gareth Nellis. 2019. *Information, Accountability, and Cumulative Learning: Lessons from Metaketa I*. Cambridge University Press.

Farbmacher, Helmut, Martin Huber, Henrika Langen, and Martin Spindler. "Causal mediation analysis with double machine learning." *arXiv preprint arXiv:2002.12710* (2020).

Gelman, Andrew, Xiao-Li. Meng, and Hal Stern. 1996. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica sinica*, 733-760.

Gelman, Andrew, Hal S. Stern, John B. Carlin, David B- Dunson, Aki Vehtari, and Donald B Rubin. 2013. *Bayesian Data Analysis*. Chapman; Hall/CRC.

Glöckner, Andreas, Susann Fiedler, and Frank Renkewitz. 2018. 'Belastbare Und Effiziente Wissenschaft'. *Psychologische Rundschau* 69 (1): 22–36. https://doi.org/10.1026/0033-3042/a000384.

Gray, Kurt. 2017. 'How to Map Theory: Reliable Methods Are Fruitless Without Rigorous Theory'. *Perspectives on Psychological Science* 12 (5): 731–41. https://doi.org/10.1177/1745691617691949.

Harrison, Glenn W. 2014. 'Cautionary Notes on the Use of Field Experiments to Address Policy Issues'. *Oxford Review of Economic Policy* 30 (4): 753–63.

Henrich, Joseph Henrich, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr and Herbert Gintis. 2004. *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*. Oxford University Press on Demand.

Huber, John D. 2017. *Exclusion by Elections: Inequality, Ethnic Identity, and Democracy*. Cambridge Studies in Comparative Politics. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316863497.

Humphreys, Macartan, and Alexandra Scacco. 2020. ‘The Aggregation Challenge’. *World Development* 127 (March): 104806. https://doi.org/10.1016/j.worlddev.2019.104806.

Hünermund, Paul, and Elias Bareinboim. 2019. ‘Causal Inference and Data-Fusion in Econometrics’. *ArXiv:1912.09104 [Econ]*, December. http://arxiv.org/abs/1912.09104.

Low, Hamish, and Costas Meghir. 2017. ‘The Use of Structural Models in Econometrics’. *Journal of Economic Perspectives* 31 (2): 33–58. https://doi.org/10.1257/jep.31.2.33.

Maniadis, Zacharias, Fabio Tufano, and John A. List. 2014. "One swallow doesn't make a summer: New evidence on anchoring effects." *American Economic Review* 104.1: 277-90.

Open Science Collaboration (OSC; (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251).

Paluck, Elizabeth Levy, Seth A. Green, and Donald P. Green. 2019. ‘The Contact Hypothesis Re-Evaluated’. *Behavioural Public Policy* 3 (02): 129–58. https://doi.org/10.1017/bpp.2018.25.

Pearl, Judea. 2009. *Causality*. Cambridge University Press.

Pearl, Judea, and Elias Bareinboim. 2011. ‘Transportability of Causal and Statistical Relations: A Formal Approach’. In *2011 IEEE 11th International Conference on Data Mining Workshops*, 540–47. Vancouver, BC, Canada: IEEE. https://doi.org/10.1109/ICDMW.2011.169.

Pettigrew, Thomas F., and Linda R. Tropp. 2006. ‘A Meta-Analytic Test of Intergroup Contact Theory’. *Journal of Personality and Social Psychology* 90 (5): 751–83. https://doi.org/10.1037/0022-3514.90.5.751.

Scacco, Alexandra, and Shana S. Warren. 2018. ‘Can Social Contact Reduce Prejudice and Discrimination? Evidence from a Field Experiment in Nigeria’. *American Political Science Review* 112 (3): 654–77. https://doi.org/10.1017/S0003055418000151.

Schnitzer, Mireille E., Russell J. Steele, Michèle Bally, and Ian Shrier. 2016. ‘A Causal Inference Approach to Network Meta-Analysis’. *Journal of Causal Inference* 4 (2). https://doi.org/10.1515/jci-2016-0014.

Wilke, Anna, and Macartan Humphreys. 2020. “Field Experiments, Theory and External Validity” in *Handbook of Research Methods in Political Science and International Relations*. R Franzese and L Curini (Eds) Oxford UP.

Zhang, Kun, Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. 2012. Kernel-based conditional independence test and application in causal discovery. arXiv preprint arXiv:1202.3775.

4 Relevance of sex, gender and/or diversity

Sex, gender, and diversity are relevant for substance and process. The primary focus of the contact studies is on differences in ethnic, religious and cultural backgrounds, but differences in effects by gender or interactions with gender are likely to be important. Our application to social contact requires researcher sensitivity to identity concerns and, critically, an ability to assess likely substantive importance of design and contextual factors for moderating effects.

Appreciation of these concerns is reflected both in staff hiring and our scientific committees. WZB hiring processes involve representation of diversity officers and require attention to diversity and representation of disabled and marginalized groups as a matter of routine. The IPI unit at WZB reflects these orientations, the unit is a diverse and vibrant work environment, employing twelve researchers (including four student assistants) from eleven countries, and counting six men and six women.

Our proposed scientific committee reflects these concerns also.

5 Supplementary information on the research context

5.1 Ethical and/or legal aspects of the project

5.1.1 General ethical aspects

Since we will be working with data from experimental interventions already done, we do not anticipate any actual or potential risks and/or harm to individuals or groups. We do not expect to require approval from IRB or ethical committees however we will submit to the WZB ethics committee to confirm this determination.

We will monitor that the studies added to the meta-analysis adhered to ethical guidelines of the discipline and have received appropriate IRB approval.

We will protect confidentiality of information at all times, working only with anonymized data that cannot permit the identification of individual respondents.

5.2 Data handling

This project will be implemented as an open access project, with anonymized data that we access made publicly available and stored on accessible repositories such as the Open Science Framework and Dataverse. Analysis code, hosted on Github, will draw from these repositories.

5.3 Other information

While the project will be hosted at WZB the project but we are seeking a close collaboration with the broader EGAP Metaketa initiative. This connection is natural since PIs Humphreys and Scacco are both member of EGAP. Further, EGAP is already looking at ways to expand their Metaketa initiative which provides plenty of room for collaboration between both groups. Lastly, the repository of harmonized studies done under the Metaketa initiative is growing faster, which provides high-quality studies to feed the Rolling Metaketa.

6 People/collaborations/funding

6.1 Employment status information

Humphreys, Macartan, Director of the Institutions and Political Inequality Unit at WZB Berlin Social Science Center, with tenure. Professor of Political Science, Columbia University, with tenure.
Scacco, Alexandra, post-doc researcher, with contract valid until 31 August 2023.
Solis Arce, Julio Saul, predoctoral researcher, with contract valid until 31 August 2021.

6.2 First-time proposal data

Humphreys, Macartan.
Scacco, Alexandra.
Solís Arce, Julio Saul.

6.3 Composition of the project group

Humphreys, Macartan, Director of the Unit Institutions and Political Inequality at the WZB Berlin Social Science Center and Professor at Columbia University, permanent contract.
Scacco, Alexandra, Senior Research Fellow at the Unit Institutions and Political Inequality at the

WZB Berlin Social Science Center, with contract in vigor until 31 August 2023.

Solis Arce, Julio Saul, Predoctoral Research Fellow at the Unit Institutions and Political Inequality at the WZB Berlin Social Science Center, with contract in vigor until 31 August 2021.

6.4 Researchers in Germany with whom you have agreed to cooperate on this project

Rolling Metaketa:

Bernd Beber (RWI - Leibniz Institute for Economic Research) will be a member of the scientific committee, contributing in the theory specification and the identification of new experimental research designs to further power the analysis of mechanisms between contact and the reduction of prejudice in the causal meta-analysis.

Ruth Dittlman (Hertie School of Governance) will be a member of the scientific committee. She will provide key expertise in the theory specification and the identification of research initiatives that could feed the meta-analysis sections of the contact Rolling Metaketa, such as the Zurich Intergroup Project.

Meta-Rep cross-project collaboration (if projects are funded):

Lennart Kaplan (Göttingen); Menusch Khadjavi (Kiel); Kacana Sipangule Khadjavi (Kiel); and Amma Panin (UCLouvain) are submitting a proposal to replicate canonical results from laboratory experiments made in Western countries by running them in developing countries (Zambia, Ghana and South Africa). Our collaboration could take the form of 1) setting a Rolling Metaketa for their project by specifying a high level theory in which economic interactions are shaped by culture; and aggregating the results from their Western and non-Western samples or 2) setting a Rolling Metaketa to aggregate results from experiments from different samples (including their non-Western replication studies but also the results from studies they are replicating) to learn about the validity of the theory motivating the experimental design. Both are promising and contribute to map explicitly how their replication attempts contribute to knowledge accumulation.

Jörg Peters (RWI - Leibniz Institute for Economic Research) and his team, Nathan Fiala and Anna Dreber, are submitting a proposal consisting in a large-scale replication (using original data) of a) IV-based papers and b) RCT-based papers and run “replication markets”. We see that both approaches can interact in a variety of ways; for instance, best guesses on the “replication market” could be derived from the transportability section of the Rolling Metaketa, be it by using covariates from a target population or simulating data. If the original replication of studies is theory based, there is clearly potential to make “replication markets” be powered by a “Rolling Metaketa”; therefore, connecting our project even more with the concept of replication

6.5 Researchers abroad with whom you have agreed to cooperate on this project

Cyrus Samii (New York University and Executive Director of EGAP) and **Jake Bowers (University of Illinois and Methods Director of EGAP)** will provide key methodological insights in the development of the aggregation tools of the Rolling Metaketa framework. Their input will provide methodological expertise, insights on experiences with past coordinated trials, and an avenue to mainstream the use of the approach if successful.

Jonas Heirman (World Food Program) will help us connect the framework of the Rolling Metaketa with the broader impact evaluation initiative from the World Food Program. In this way, will seek insights to help ensure that our framework is helpful to guide policy design and implementation.

Claire L. Adida (University of California, San Diego) will be part of the scientific committee providing expertise from her work on migrant integration relevant for theory specification and the selection of studies included in the contact Rolling Metaketa.

Rebecca Wolfe (University of Chicago) will be part of the scientific committee providing expertise from her work in peace-building and post-conflict recovery for theory specification and the selection of studies included in the contact Rolling Metaketa.

Anna Wilke (Columbia University and EGAP) and Jaclyn Leaver (EGAP) are leading initiatives at EGAP to explore different strategies to develop the Metaketa model. Working with them will strengthen the development of the aggregation methods and the user interface of the Rolling Metaketa; help us build on lessons learned from previous studies done under EGAP's Metaketa initiative, and facilitate take-up if this model is deemed successful.

6.6 Researchers with whom you have collaborated scientifically within the past three years

Bernd Beber, Graeme Blair, Jasper Cooper, Alexander Coppock, Alberto Diaz Cayeros, Kim Dionne, Phillip Dawid, Thad Dunning, Miriam Golden, Guy Grossman, Susan Hyde, Alan Jacobs, Dean Karlan, Rebecca Littman, Craig MacIntosh, Edward Miguel, Salma Mousa, Mushfiq Mubarak, Monica Musio, Garreth Nellis, Carlo Prato, Alexandra Siegel, Maarten Voors, Shana Warren, Chagai Weiss, Anna Wilke, Dean Yang.

6.7 Project-relevant cooperation with commercial enterprises

If applicable, please note the EU guidelines on state aid or contact your research institution in this regard.

None.

6.8 Project-relevant participation in commercial enterprises

Information on connections between the project and the production branch of the enterprise

None.

6.9 Scientific equipment

List larger instruments that will be available to you for the project. These may include large computer facilities if computing capacity will be needed.

WZB Computer cluster for computation and hosting of shiny applications.

6.10 Other submissions

List any funding proposals for this project and/or major instrumentation previously submitted to a third party.

In 2019, Humphreys and Scacco together with Claire Adida, Bernd Beber, Ruth Dittlman, and Rebecca Wolfe submitted an unsuccessful off cycle application for an "Off the Beaten Track" grant from Volkswagen foundation. The proposal focused on the implementation of a new set of social contact experiments and not on the creation of a Rolling Metaketa.

7 Requested modules/funds

7.1 Basic Module **321,510 €**

7.1.1 Funding for staff **257,800 €**

1 Postdoctoral researcher, TVöD E14 or E13, 100%, 36 months 222,300 €

- Lead, together with PI Scacco, the implementation of the first Rolling Metaketa: Intergroup contact.
- Research and develop the necessary methods of inference aggregation used in the causal meta-analysis and the transportability section of the Rolling Metaketa.
- Conduct the literature review previous to a meta-analysis.
- Communicate with scholars and retrieve data from primary studies to set the meta-analysis.
- Coordinate with technical staff in setting up the user interface of the Rolling Metaketa.
- Consult and collaborate with the scientific committee to refine the inference aggregation methods.
- Contribute to analysis and drafting of research products.
-

Support by student assistant (TV Stud III Berlin, 15 hours/week) for 36 months, 35,500 €

- Assist with the gathering of data from original studies.
- Replicate previous meta-analysis.
- Data cleaning.
- Assist in the programing of the user interface of the Rolling Metaketa.

7.1.2 Direct Project Costs **42,540 €**

7.1.2.2 Travel (calculated acc. to BRKG)

Travel cost are requested for travel to international scientific conferences for the presentation of the project (6 trips).

Travel to conferences: 3 trips to conferences in Europe (3 * 1,200€), 3 trips to conferences in non-European countries (3 * 1,900€) 9,300 €

- American Political Science Association annual meeting; European Political Science Association annual meeting; EGAP annual meeting; Causal Data Science Meeting.

7.1.2.5 Other Costs

18,000 €

External Expertise for project related subjects (6 * 3,000 Euro)

- Text for opening workshop in methods and the literature on social contact

Research stays for guest researchers in Berlin (travel costs and accommodation/per diem), 3 * 2 months — 12,990 €

- Summer visits from members of the scientific committee and the core team to Berlin in order to contribute to the advances of the project and to present their own research at WZB; with the other Meta-Rep projects; and other German institutions.

7.1.2.6 Project-related publication expenses

2,250 €

Lump sum, 750 € per year

7.6 Module Workshop Funding — 21,170 €

We apply for funding for two workshops in Berlin: 1) Kick-off workshop for project team and invited experts and 2) Final workshop for project team and invited experts for dissemination of findings.

Kick-off workshop: 2 days, 25 participants, calculated are travel expenses, accommodation for collaborators and experts (2 x Germany, 3 x European countries, 10 x non-European countries), accommodation, daily allowances, catering — 21,170 €

Furthermore, a 22% overhead contribution to cover indirect costs is requested.