

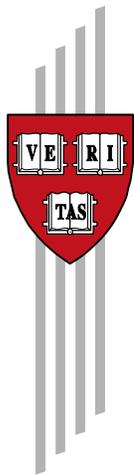
When Do Development Projects Enhance Community Well-Being?

Michael Woolcock

CID Faculty Working Paper No. 355

June 2019

© Copyright 2019 Woolcock, Michael; and the President and Fellows of
Harvard College



Working Papers

Center for International Development
at Harvard University

When do development projects enhance community well-being?

Michael Woolcock¹

June 2019

Abstract

Many development agencies and governments now seek to engage directly with local communities, whether as a means to the realization of more familiar goals (infrastructure, healthcare, education) or as an end in itself (promoting greater inclusion, participation, well-being). These same agencies and governments, however, are also under increasing pressure to formally demonstrate that their actions ‘work’ and achieve their goals within relatively short timeframes – expectations which are, for the most part, necessary and desirable. But adequately assessing ‘community-driven’ approaches to development requires the deployment of theory and methods that accommodate their distinctive characteristics: building bridges is a qualitatively different task to building the rule of law and empowering minorities. Moreover, the ‘lessons’ inferred from average treatment effects derived from even the most rigorous assessments of community-driven interventions are likely to translate poorly to different contexts and scales of operation. Some guidance for anticipating and managing these conundrums are provided.

Introduction

One of the many sea-changes in the field of international development in recent decades has been the expansion of financial resources, political support, advocacy efforts and scholarly activity afforded to ‘communities’, as both the explicit means and ends of programmatic interventions. These changes reflect an underlying reassessment, spanning roughly seven decades, of the role of social institutions in the development process (Greif and Iyigun 2013, Woolcock 2017). In the early 1950s, for example, the United Nations (1951:15) could publish a flagship document asserting that “...rapid economic progress is impossible without painful adjustments. Ancient philosophies have to be scrapped: old social institutions have to disintegrate; bonds of caste, creed and race have to burst; and large numbers of people who cannot keep up with progress have to have their expectations of a comfortable life frustrated. Very few communities are willing to pay the full price of economic progress.”² From a development standpoint, in short, prevailing social institutions were a problem, a real and present obstacle needing to be ‘burst’ if ‘progress’ was to be made.

Today, is it unthinkable that international agencies could make such statements (at least in their public documents). If in the 1980s the status of social institutions ‘improved’ to being merely epiphenomenal – i.e., subservient to more fundamental economic and political forces – by the early 1990s Robert Putnam

¹ Development Research Group, World Bank, and Kennedy School of Government, Harvard University. An earlier version of this paper was presented as an opening address to Evaluation Week 2019, an annual event hosted by the World Bank’s Independent Evaluation Group. My thanks to Howard White, Alison Evans, Elliot Stern and conference attendees for helpful questions and comments. The views expressed in this paper are those of the author alone, and should not be attributed to the World Bank, its executive directors or the countries they represent. Contact information for correspondence: michael_woolcock@hks.harvard.edu.

² Cited in Escobar (1995: 3).

(1993) could famously declare that certain configurations of civic life were in fact central to “making democracy work”. Twenty-five years on, the World Bank oversees an entire portfolio of national projects and analytical work known collectively as ‘Community Driven Development’ (hereafter CDD), supporting more than 190 such projects in 78 countries (Guggenheim and Wong 2018). These activities, and especially the major projects to which they give rise, seek to attain not only familiar objectives such as raising the physical well-being of participants (e.g., in the form of enhanced incomes, crop yields and water security) and connecting them to markets (via small-scale roads and bridges) but, more ambitiously, to promote greater social inclusion in community deliberations and decision-making, to enhance the quality and cost-effectiveness of service delivery (e.g., schools, health, finance, justice), and even to foster a more coherent and legitimate ‘social contract’ between citizens and the state. Flagship CDD projects have been implemented in countries as challenging as Afghanistan and Myanmar and as diverse as Indonesia and Nigeria.³

In principle, of course, this sounds like a genuine advance in the way development is conceived and undertaken. If the policies and strategies of multilateral agencies such as the World Bank were once (and in certain circles continue to be) criticized for their (alleged) indifference to social relations (e.g., flooding sacred valleys to create massive hydroelectricity dams), for being thin cover for a ‘neo-liberal’ reform agenda, or for being an instrument of Western foreign policy, who could now be against an approach overtly promoting greater participation of women and the poor in village decisions, encouraging greater transparency and accountability for how public resources are allocated, or promoting greater social cohesion and more effective state-society relations?

Well, quite a lot of people, it turns out, but for rather different reasons, most of which become apparent in the ways supporters and critics make claims (and counterclaims) regarding the efficacy of CDD. In this short paper, I focus on the key concerns that have emerged as researchers and evaluators have sought to answer the seemingly reasonable question, ‘Do community-based approaches to development work?’ I will argue that this question, as it is usually presented and understood, is well-meaning but too often misleading, less because of the quality of evidence collected than the basis on which inferences are drawn from it. CDD-type interventions – unlike more familiar development interventions such as childhood immunizations, agricultural subsidies and nutrition supplements – fundamentally do not, and inherently cannot, yield uniform (and uniformly predictable) outcomes; they have distinctive characteristics (outlined below) that preclude them from doing so. This of course does not absolve CDD projects and project staff of being held to account, of their efforts being assessed, and of broader conclusions needing to be drawn about the likely effectiveness (on average) for a given CDD project in new places or at larger scales of operation; it does mean that the distinctive characteristics of CDD interventions requires them to be evaluated, and the findings from these evaluations interpreted, in

³ Needless to say, the idea and practice of incorporating ‘communities’ into development programming itself has a long history, much of it emanating from poverty analysis and agricultural extension (though obviously space constraints prohibit such a discussion here). The key features of what the World Bank and others today call community-driven development is the disbursement of block grants to districts, which then allocate funds to community groups who, within specified rules (e.g., money can’t be used for religious purposes), propose a small project (a bridge, meeting house), the merits of which are assessed by their peers and voted upon. The sum total of the value of the proposed projects exceeds the amount allocated, so some proposals ‘win’ and others do not – needless to say, this creates tension, which needs to be anticipated, accommodated and addressed. The rules determining the composition of each group often specify that at least one member should be a woman, and that the meeting to allocate the funds cannot be held until a proposal is received by a women-majority group. (For an overview of the core features of CDD programs from the World Bank’s perspective, see <https://www.worldbank.org/en/topic/communitydrivendevelopment>)

ways commensurate with this distinctiveness. A better question – which is to say, a more fruitful and useable question – is: ‘*When* do development projects enhance community well-being?’

Why Assessing and Drawing Inferences from CDD Interventions is so (Inherently) Difficult

It is hard to design development interventions, of any kind, and harder still to implement them. But once this substantive work is done, sooner or later the time will come for an ‘evaluation’ (of one form or another) to be conducted on this intervention to determine whether it has ‘worked’. For the least sophisticated (or cash-strapped) development interventions, a few cursory interviews with program staff and participants may be held, along with an inspection of accounting records; if everything appears to be ‘in order’ some photographs of smiling faces will be taken, a few happy stories collected, and a short report duly issued. For the most sophisticated and well financed interventions, the evaluation moment will have been not merely anticipated but carefully planned for during the design phase. Some combination of limited resources, policy objectives (such as poverty reduction) and political imperatives usually means that most local-level development interventions cannot be provided to everyone, necessitating some form of demographic (specific people) and/or geographic (specific places) targeting. For astute evaluators, such requirements present a challenge and an opportunity to carefully identify matched – or, better yet, randomly assigned (within the necessary demographic and geographic parameters) – groups of program participants (a ‘treatment’ group) and otherwise identical non-participants (a ‘control’ group) from whom baseline and follow-up data on key outcome indicators can be collected. By then calculating a ‘double difference’ – between treatment and control groups, and their corresponding baseline and follow-up scores on the key indicators – an average treatment effect across these indicators is thereby discerned. If the resulting numbers are net positive, the program is deemed to have ‘worked’.⁴ The more ‘rigorous’ the methodological strategy deployed, the higher the confidence one has in this conclusion and thus the broader inference(s) drawn from it.

This is the core underlying logic that informs how most serious local-level⁵ development evaluations are conducted. But, I will now argue, conclusions drawn on the basis of neither of these options – the least nor allegedly most sophisticated evaluation protocol – provides sufficient grounds on which to declare that a given CDD intervention has or (especially) has not ‘worked’. Why? The limits of the least sophisticated approach should be readily apparent to anyone with solid methodological training in the social sciences: all sorts of factors, known (selection bias) and unknown (unobserved variation in either treatment or control groups, such as motivation and leadership quality), could be driving the observed outcome(s), and the tiny (non-random, non-representative) “sample” on which any claims about program-wide impact are being made simply cannot accommodate (or enable the evaluator to ‘control

⁴ An array of evaluation protocols – e.g., natural experiments, quasi-experiments, regression continuity designs, propensity score matching – can be deployed to elicit the necessary counterfactual group (i.e., an otherwise comparable population who did not participate in the program whose outcome variables of interest can also be tracked over time). Recent pragmatic innovations in qualitative methodology (e.g., Copestake et al 2019) enable organizations operating at modest scale with humble budgets to generate useful (‘good-enough’) information on how their intervention is working; if not ideal, these approaches nonetheless constitute a significant and welcome advance (see also Bamberger, Vaessen and Raimondo 2016). Researchers also widely recognize that ethical, logistical, and political reasons may preclude the creation of formal treatment and control groups. In their own way, recent econometric advances have also enabled evaluators (at least of large social programs) to estimate ‘synthetic controls’ – i.e., to compute by careful extrapolation from existing data what a control group’s baseline and follow-up characteristics would look like if in fact there was one (see Athey and Imbens 2017).

⁵ By ‘local’ I mean the primary unit of analysis at which the development intervention is targeted; accordingly, raising interest rates and building shipping ports is not local (or ‘micro’ or ‘social’); seeking to enhance learning outcomes in schools or reducing the incidence of poverty via cash transfers (usually) is.

for') any of these confounding biases and contending influences. Claims emerging from such 'evaluations' are typically (and rightly) dismissed as being "merely anecdotal".

The limits of the most sophisticated evaluation strategies, however, are less obvious but deeply problematic nonetheless. The power of a randomization protocol, for example, relies on a core assumption, namely that each member of the 'treatment' group receives the same type, frequency and duration of the intervention.⁶ Canonically, this is how new medicines are tested: participants must receive exactly the same pills, take the same number of pills at the same time each day, take no other pills, and stay with the program (and adopt consistent behavioral and dietary patterns) for the same length of time. (They should also not be aware of whether they are in the treatment or control group, and neither should the person who subsequently conducts the data analysis.) But participants in a CDD program – and there can be millions of them, as in Indonesia – are really not receiving an 'identical' treatment. *By design*, a central task of frontline implementers of such programs is to adhere closely to implementation rules while nonetheless making adjustments to accommodate idiosyncratic contextual realities, few of which could ever be fully anticipated and specified during the program's design phase. Enhancing the effectiveness with which such judgments are made by facilitators – in real time, under pressure, across sub-national contexts with anthropological levels of diversity – can only partially be taught, formally assessed and uniformly enforced; which is to say, the facilitators' effectiveness will vary enormously. Taken together, each group in a CDD program can experience the ostensibly 'same' program in very different ways, where that difference is a function of unique intra-group dynamics, the nature of the group's relationship with their facilitator, the facilitator's skill, and the array of political, economic and anthropological factors that together constitute the 'context' within which all these interactions take place.

The implications of these realities become most consequential when a null or negative verdict is rendered on a particular CDD program. If a net positive outcome has been attained, the broader (known) confounding factors having been duly accommodated in the evaluation design and subsequent analysis, then one can reasonably assume that this outcome has transpired *despite* all the forms, degrees and sources of variation and the associated implementation challenges.⁷ This same array of variation, however, in the event of a net zero or negative impact verdict being initially rendered, makes it premature to conclude that the CDD program in question "didn't work", precisely because it remains unclear why and what it was, exactly, that did not work. Was an otherwise solid design poorly implemented? Was a weak design implemented by a highly capable team, thus meaning that an *even worse* outcome might have been attained had more modestly skilled implementers been assigned to the task? Was a solid design and diligent implementation undone by a political context that just happened to be deeply unfavorable to this type of intervention? Null results are under-represented in the scholarly literature, but Rao, Ananthpur and Malik (2017) unpack such a result from a livelihoods project in India, showing that while the *average* treatment effect was zero there was nonetheless high variation around this mean: the project actually worked fine for some groups in some places even as it also made other groups in other places worse off, raising the obvious question: under what conditions did this intervention 'work', under what conditions did the same intervention have no impact, and under what conditions did it utterly fail? In short, *when* did it work or not? This question, I suggest, is the more

⁶ This is known, formally, as the unit homogeneity assumption.

⁷ At least within the timeframe over which the evaluation was conducted. As I argue below, the (high) likelihood of CDD projects having a non-linear impact trajectory over time means positive initial impacts could steadily taper and even dramatically decline in subsequent periods.

fruitful one to ask of community-based interventions, precisely because basic theory and experience strongly suggests that high variance in the outcome space is to be routinely expected.

Broader Considerations

Compounding these inference challenges regarding the efficacy of CDD-type interventions is the even more vexing concern that variation is likely to transpire not only across space and groups, but time. Non-linear, non-monotonic impact trajectories are highly likely to characterize large CDD interventions (Barron, Diprose and Woolcock 2011) and most likely a host of other development interventions as well (Woolcock 2009), meaning that any claim about impact should be conditional on (a) when the follow-up data was collected and (b) where this data collection point sits in the predicted (on the basis of theory, evidence and experience) impact trajectory one would reasonably associate with this particular intervention implemented in this particular place by this particular team for this particular group.

Inadequate consideration of such issues leads to type of summary conclusions reached in Casey, Glennerster and Miguel (2012): their methodologically exemplary impact evaluation of a CDD program in Sierra Leone seeking to simultaneously create jobs and improve community cohesion found “positive short-run effects on local public goods and economic outcomes, but no evidence for sustained impacts on collective action, decision making, or the involvement of marginalized groups, *suggesting that the intervention did not durably reshape local institutions*” (p. 1755, emphasis added). This latter inference can be entirely supported by ‘the evidence’ yet be unwarranted if, as I think social theory would suggest is highly likely, the underlying impact trajectories of the economic and social components have qualitatively different ‘shapes’. In its simplest form, what conclusions would one draw if an otherwise methodologically rigorous evaluation of this intervention was conducted three years after launch, but where the expected *economic* change trajectory is linear and uniformly increasing over time while the expected *social* change trajectory is essentially flat for (say) five years before slowly but steadily rising? One would likely conclude that the economic components did indeed ‘work’ but that it was too soon (by two years) to make a call on the efficacy of the social components. More precisely, however, one should conclude that it is unclear whether the social components of this project are doing fine and should just stay the course, or whether they are actually flailing and will never generate the desired outcome or impact, or whether the null result is actually welcome progress (because, *in this instance*, the change trajectory turns out to be a ‘j curve’ – i.e., getting worse before it eventually, maybe, gets better – and in the hands of less able implementers would otherwise, at this point, be doing considerable harm).⁸

To distinguish between these different possibilities – and doing so surely matters for operational and ethical reasons – requires the incorporation of different types of research methods and theory. A singular methodology, no matter how putatively ‘rigorous’, cannot yield correct inferences without reference to a credible theory of change that provides guidance as to what it is reasonable to expect by when. And if, as I suggested above, there is also likely to be wide geographic and demographic variation in outcomes in CDD projects – actually, complex interventions of all types – then a sound evaluation strategy will need to be accompanied by a comprehensive fit-for-purpose monitoring strategy to enhance the quality of implementation and adjustments to shifting contextual realities. If there has been a veritable ‘revolution’ in development evaluation in recent decades, strongly pushed by researchers, the emerging profile of socially complex interventions (such as CDD), and the ensuing debates surrounding their effectiveness, strongly suggests the need for corresponding complementary efforts to ‘upgrade’ the purpose, status and sophistication of monitoring, shifting it from being an

⁸ Such a trajectory shape has been observed in women’s empowerment projects (in which men initially react to their newly assertive wives by suppressing them even more) and in governance reforms (in which entrenched incumbents resort to violence to resist citizens now demanding greater transparency and accountability).

onerous under-funded instrument of compliance to a valued and useable tool for eliciting real-time feedback and organizational learning.

One important implication of these issues is that ‘systematic reviews’ of evaluations conducted on the broad category of ‘community-driven development’ interventions – i.e., meta-analyses of the most sophisticated individual studies undertaken in a particular field, designed to elicit conclusions regarding the overall effectiveness of this class of interventions – must be handled with great caution. Such reviews make sense in fields where there is a clear understanding of the mechanisms connecting inputs and impacts, and where these mechanisms – behavioral, physiological, pharmacological, economic – essentially function the same way independently of context and where implementing the intervention itself requires relatively low organizational capability (e.g., dispensing vitamin supplements versus consolidating peace agreements). Put more formally, systematic reviews work best for interventions with low causal density⁹ and high external validity, while grounded case studies are more suited to interventions with high causal density and low external validity (Woolcock 2013).

As I hope the preceding discussion has shown, however, CDD-type interventions rarely meet these two criteria: they are replete with an array of complex social mechanisms that are explicitly designed to accommodate contextual idiosyncrasies, and implementing them all – coherently, consistently, legitimately, effectively, at scale – requires organizations with high levels of capability. The wide variation in outcomes that can be expected of CDD-type interventions, even under the most favorable circumstances, means that universal claims about whether they generically ‘work’ (or not) should be replaced by a focus, in each particular case, on identifying the conditions under which this diversity of outcomes – for different groups in different places at different times – is experienced.

For such interventions, the more fruitful methodological and empirical quest should be to ‘look beyond averages’ (cf. Ravallion 2001) to identify the “key facts” (Cartwright and Hardie 2012: 137; see also Woolcock 2013) driving this heterogeneity in each case. Doing so requires deploying mixed methods strategies, in particular those capable of discerning the ‘causes of effects’ and not just the ‘effects of causes’¹⁰ (see Woolcock 2019). Systematic reviews and meta-analyses of CDD interventions (White, Menon and Waddington 2018, Casey 2018) are not antithetical to this quest; they can and have yielded insights about which researchers, practitioners and potential adopters should be aware as efforts are made to adapt and improve core designs (e.g., to urban settings, where CDD interventions have often struggled, and via building on/with ‘organic’ social institutions rather than ones ‘induced’ by external agents¹¹). But learning that the portfolio of CDD projects is, one average, better at producing local infrastructure than fostering more inclusive or ‘empowered’ social institutions is entirely what an ounce of social theory would predict; in this domain, systematic reviews should not be the final or only arbiter of how policy conclusions are drawn. Enhancing community well-being is more social work than clinical medicine, and no less important or challenging for being so. It should be undertaken as a necessary complement to, not an alternative substitute for, broader development strategies to forge an effective, accountable and responsive government.

Conclusions

Researchers and administrators of development interventions, of all types, face increasing pressure to demonstrate the utility of their efforts, to show that finite resources are being optimally deployed and

⁹ Drawn from neuroscience, computing and physics, the concept of ‘causal density’ refers to the number of independent interactions occurring within a particular system (see Manzi 2012).

¹⁰ This deft distinction was initially made by John Stuart Mill; see also Goertz and Mahoney (2012).

¹¹ This is a key conclusion of the important review conducted by Mansuri and Rao (2012).

are meeting stated objectives (on time, on budget). For the most part, the spirit of these expectations is desirable and meeting them is possible. But realizing even a slice of the vastly expanded scale and scope of the contemporary development agenda – as manifest in the Sustainable Development Goals, ratified by 193 countries, with its 157 targets spanning poverty elimination and lifelong learning to climate change and effective institutions ‘at all levels’ – surely requires calling upon the full arsenal of social science theory, methods and inference. Nowhere should this imperative be more prescient than in efforts to assess the veracity of efforts to expand the nature and extent of the role played by ‘communities’ in these processes, whether as an end in itself or a means to building cost-effective infrastructure or forging productive state-society relations.

Despite strong prevailing expectations and imperatives to show whether community-driven development projects “work” in some generic and/or singular sense, I argue instead that we should begin from a premise that development is disruptive, by design, and that technically sound design features are necessary but not sufficient to ensure that community-driven development initiatives achieve their desired outcomes. Precisely because ‘communities’ are both the means and ends of such interventions, vastly more attention needs to be given to understanding (and accommodating) the inherently wide array of outcomes that they are likely to generate. This variance will be a function of deep contextual idiosyncrasies (including local political dynamics), high variation in implementation capabilities, the (perceived) legitimacy of the change process, and the strong likelihood that, even in the most fortuitous of circumstances, impact trajectories over time will be highly non-linear. These realities have important implications for the *kinds* of expectations, empirical claims and policy inferences we make of all development interventions, but especially those seeking explicitly to enhance community well-being, in high and low income countries alike. To get good answers to these questions, we need to ask when in particular, not whether in general, development projects to enhance community well-being ‘work’.

References

- Athey, Susan and Guido Imbens (2017) ‘The State of Applied Econometrics: Causality and Policy Evaluation’ *Journal of Economic Perspectives* 31(2): 3-32
- Bamberger, Michael, Jos Vaessen and Estelle Raimondo (eds.) (2016) *Dealing with Complexity in Development Evaluation: A Practical Approach* Los Angeles: Sage Publications
- Barron, Patrick, Rachael Diprose and Michael Woolcock (2011) *Contesting Development: Participatory Projects and Local Conflict Dynamics in Indonesia* New Haven: Yale University Press
- Cartwright, Nancy and Jeremy Hardie (2012) *Evidence-Based Policy: A Practical Guide to Doing it Better* New York: Oxford University Press
- Casey, Katherine (2018) ‘Radical Decentralization: Does Community-Driven Development Work?’ *Annual Review of Economics* 10: 139-163
- Casey, Katherine, Rachel Glennerster and Edward Miguel (2012) ‘Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan’ *Quarterly Journal of Economics* 127(4): 1755-1812
- Copestake, James, Marlies Morsink and Fiona Remnant (2019) *Attributing Development Impact: The Qualitative Impact Protocol Casebook* Rugby, UK: Practical Action Publishing

Escobar, Arturo (1995) *Encountering Development: The Making and Unmaking of the Third World* Princeton, NJ: Princeton University Press

Goertz, Gary and James Mahoney (2012) *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences* Princeton, NJ: Princeton University Press

Greif, Avner, and Murat Iyigun (2013) 'Social Organization, Violence, and Modern Growth' *American Economic Review: Papers & Proceedings* 103(3): 534–538

Mansuri, Ghazala and Vijayendra Rao (2012) *Localizing Development: Does Participation Work?* Washington, DC: World Bank

Manzi, Jim (2012) *Uncontrolled: The Surprising Payoff of Trial and Error for Business, Politics, and Society* New York: Basic Books

Putnam, Robert (1993) *Making Democracy Work: Civic Traditions in Modern Italy* Princeton, NJ: Princeton University Press

Rao, Vijayendra, Kripa Ananthpur and Kabir Malik (2017) 'The Anatomy of Failure: An Ethnography of a Randomized Trial to Deepen Democracy in Rural India' *World Development* 99(11): 481-497

Ravallion, Martin (2001) 'Growth, Inequality and Poverty: Looking Beyond Averages' *World Development* 29(11): 1803-1815

United Nations (1951) *Measures for the Economic Development of Underdeveloped Countries* New York: Department of Social and Economic Affairs, United Nations

White, Howard, Radhika Menon and Hugh Waddington (2018) 'Community-Driven Development: Does it Build Social Cohesion or Infrastructure? A Mixed-Method Evidence Synthesis' Technical Report. New Delhi: International Initiative for Impact Evaluation (3ie)

Wong, Susan and Scott Guggenheim (2018) 'Community-Driven Development: Myths and Realities' Policy Research Working Paper No. 8435. Washington, DC: World Bank

Woolcock, Michael (2009) 'Toward a Plurality of Methods in Project Evaluation: A Contextualized Approach to Understanding Impact Trajectories and Efficacy' *Journal of Development Effectiveness* 1(1): 1-14

Woolcock, Michael (2013) 'Using Case Studies to Assess the External Validity of Complex Development Interventions' *Evaluation* 19(3): 229-248

Woolcock, Michael (2017) 'Social Institutions and the Development Process: Using Cross-Disciplinary Insights to Build an Alternative Aid Architecture' *Polymath: An Interdisciplinary Arts and Sciences Journal* 7(2): 5-30

Woolcock, Michael (2019) 'Reasons for Using Mixed Methods in the Evaluation of Complex Projects', in Michiru Nagatsu and Attilia Ruzzene (eds.) *Contemporary Philosophy and Social Science: An Interdisciplinary Dialogue* London: Bloomsbury Academic, pp. 149-171