

Organizational Mechanisms of Scaling Social Enterprises

Christian Seelos¹ and Johanna Mair

Christian Seelos

Visiting Scholar
Center on Philanthropy and Civil Society
Stanford University
Encina Hall, E 106
616 Serra Street
Stanford CA 94305
Email: cseelos@stanford.edu

Johanna Mair

Hewlett Foundation Visiting Scholar
Academic Editor Stanford Social Innovation Review
Center on Philanthropy and Civil Society
Stanford University
Encina Hall, E 106
616 Serra Street
Stanford CA 94305
Email: jmair@stanford.edu

¹ Corresponding Author

Organizational Mechanisms of Scaling Social Enterprises

Christian Seelos and Johanna Mair

Abstract

In the social sector, the subject of scaling has recently emerged as a dominant discourse. The scholarly literature on this phenomenon is in an embryonic stage. In this paper, we provide a clear definition of scaling that makes it a subject for mechanism-based scholarly investigation. An analytical model of organizational scaling is developed that is grounded in a critical realist philosophy of science. Critical realism provides a sophisticated view of reality that enables explicit operationalization of the causal links between actors, mechanisms and outcomes. The analytical model thus provides guidance to and structure for the empirical investigation of an extraordinary organizational example of scaling. A number of propositions for scaling are developed and we reflect on the implications for practitioners and further scholarly research.

Much Ado about Scale. The buzz word in social enterprise is “scale.” This is the title of an article by Adrenne Villani (2010) in “Beyond Profit”, an online magazine that focuses on social enterprises that target the social challenges of poverty. The author identifies the topics of scale and scaling as a key discourse amongst individuals and organizations engaged in the fight against poverty. This trend was also recognized by scholars. In their paper *Think Large and Act Small: Toward a New Paradigm for NGO Scaling Up*, Uvin et al. (2000) attend to scaling as a key topic in the social sector. Scholarly literature on this phenomenon is only slowly emerging and has not yet generated a clearly defined set of concepts. We do not have a consistent set of perspectives on what the phenomenon of scaling actually is or how to study it. It is the main objective of this paper to clarify the concept of scaling in the social sector and develop ways to move scaling from a practitioner discourse to make it an object of systematic scholarly investigation. This is our strategy: we engage with the recent call for more mechanism-based explanations in the social sciences (Hedström and Swedberg, 1998; Mahoney, 2001; Mayntz, 2004; Demetriou, 2009; Hedström and Ylikoski, 2010). Mechanism based explanations require a clear explanandum as a starting point for inquiry (Mayntz, 2004). To facilitate this we first develop a definition of scaling that makes it a subject for mechanism-based scholarly investigation. We then develop an analytical model of organizational scaling to provide guidance to and structure for our empirical investigation of an extraordinary organizational example of scaling. Our analytical model is grounded in a critical realist philosophy of science because of its sophisticated specification of the structure of reality that links actors, mechanisms and outcomes.

Scaling – discourse and activities

Scaling has dominated for some time now the conversations in the social sector. Two important developments drive this discourse. First, decades of investments of significant economic and development resources have concentrated 80% of global wealth in the hands of less than 10% of the world population (Davies, Sandström, Shorrocks and Wolff, 2008). One of the important consequences of this development is that inequality and poverty persist in important and intolerable dimensions (Chen and Ravallion, 2007; Easterly, 2002). Tackling poverty therefore not only requires innovative approaches and initiatives but also finding ways to size these initiatives up to meet the scale of the problem. Second, some social enterprises such as BRAC or the Grameen Bank in Bangladesh have achieved tremendous scale. A look at their annual reports indicates that they operate in more than 60000 villages and serve or productively engage millions of poor people in a variety of activities ranging from microfinance to health-services. With the proof of concept on the table this now becomes the benchmark for other organizations in the eyes of funders, advisers and policy makers.

This attention to scaling reflects both a sense of urgency and a sense of excitement and hope in the face of the grim reality of global poverty. Urgency, the availability of philanthropic financial resources and hope have stimulated a variety of scaling activities in the social sector. Some actors have found new ways to expand the meaning of scaling by telling richer stories about social enterprises and related phenomena. The benefit of this move is that social construction of a new, more exciting reality can be done with little effort and in very short time. These stories create attention, inspiration and even hope. This move is perhaps best reflected in a quote by Uvin et al. (2000, p.1418): *"In the new paradigm, the extent to which an NGO successfully scales up can be judged not only in terms of its size, but also in terms of the number of spin-offs it created, the number of projects that have been taken over by other actors, and the degree to which it contributed to the social and intellectual diversity of civil society. [...] Impact, finally, is not only about the number of beneficiaries or even the specific policy changes won, but also about local capacity built, intersectoral contacts developed, norms of trust and cooperation strengthened, and democratic space and social diversity reinforced."* Expanding the stories about existing organizations thus resonates deeply with the sense of urgency and need for hope in the social sector.

Another consequence of the focus on scaling are the many efforts to develop better measures and evidence for the outputs, outcomes and impact of initiatives. Ebrahim and Rangan (2010) have highlighted the growing dominance of this discourse in the social sector: *"A more recent manifestation of this discourse has centered on the mantra of "impact," or demonstrating results in addressing complex social problems such as poverty and inequality."* To some extent this also changes the discourse around existing organizations by demonstrating more clearly and perhaps in a much richer and complete way the real scale of their activities and areas of impact. One consequence of this is a fascination with success stories and framing the inability to scale as a sign of failure. This is reinforced by recent scholarly work that attempts to identify more

systematically the success factors that explain the success of some social enterprises. For example, Bloom and Chatterji (2009) proposed the SCALERS model which posits that an organization's success at scaling social impact will be a consequence of its capabilities in seven areas: staffing, communicating, alliance-building, lobbying, earnings-generation, replicating, and stimulating market forces. The expectation is that getting these factors right would be the fastest and most efficient way to achieve scale consistent with the needs for speedy progress and creating hope and excitement. We will argue below that this focus on success stories and success factors stifles learning and progress in developing a systematic base of scholarly knowledge on the issue of scaling in the social sector. We will show that explaining the ways in which failure can be prevented by engaging in counterfactual analysis may be more insightful than searching for success factors.

Where the discourse of scaling meets the reality of poverty

Unfortunately, current evidence does not support the hope and excitement generated by scaling efforts. One problem is that some models don't actually create the type of impact that was claimed. Recent analyses of microfinance models voiced concerns over a lack of evidence for their ability to create impact. Another problem seems to be a dependency of successful models on a particular context. BRAC, one of the largest and most successful and experienced social enterprises is struggling with initiatives outside of Bangladesh. Scholars have voiced concerns that many innovative and entrepreneurial organizations seem unable to scale. For example Uvin (1995: p. 927) states: *"It is often recognised that most of these grassroots initiatives are small, underfunded, poorly staffed, slow and localised in the face of poverty, hunger and degradation on a vast scale. They are thus considered to be only actions at the margins, capable of providing local relief and empowerment, but not of tackling the real issues of the eradication of hunger and poverty for hundreds of millions of people."* Bradach (2003) expressed similar concerns about the fragmentation, smallness and insufficient realization of the potential of proven programs.

We use the term scaling in this paper to refer to organizational activities that achieve an explicit set of empirically observable outcomes that grows over time in a meaningful and measurable dimension. Scaling as used in this paper thus implies an understanding of relevant cause/effect relationships. In operational terms, our definition of scaling implies that an organization generates outcome event regularities over significant time, i.e. that it is able to enact a set of mechanisms that generate a set of desired outcomes. In the simplest definition, we thus define scaling as an organizational phenomenon whereby doing more of A or doing A better (set of mechanisms) generates more of B (set of outcomes). However, many scholars have voiced concerns over the expectations on "social engineering" as implied by this perspective on scaling. Merton has highlighted the "unintended consequences" of purposeful action as an important reminder that *"social life is not as simple as it first seems"* (Merton, 1968: p.122). Contrary to scholarly efforts of searching for success factors, i.e. those factors suspected to generate intended consequences, Merton states that the discovery of unintended consequences and the circumstances of their workings represents an even more important approach for significant

progress in sociological knowledge (Merton, 1968). Portes (2000) also makes a strong statement for sociology as analysis of the unexpected. He identifies the gaps between “received theory and unexpected reality” as a main cause of skepticism in sociology and warns about ignoring “derailing factors” in favor of simplistic linear means-ends perspectives. Portes (2000) views this attention to the unexpected as an important practice of bashing myths. One of these myths implied in the discourse around scaling lies in our ability to deliberately design organizations that generate expected outcomes. But this “clashed inevitably with the paradoxes of social life”. (Portes, 2000: p 12). Charles Tilly, considered a pioneer in making advances towards mechanism-based sociological explanations (Demetriou, 2009) expands this focus on the unexpected by the perspective of including that which did not occur. “*Sound social science concerns counterfactuals: explaining what actually occurs, which ironically requires specifying what did not occur but could have occurred, then comparing factual with counterfactual.*” (Tilly, 1996: p 596). Tilly calls for much more attention to errors, their consequences and rectification. In other words social complexity means that explanatory factors lie in the factual world but potentially even more in the realm of the counterfactual. However, counterfactuals do not show up in traditional empirical factor analyses and tend to escape investigation.

The implications for our studies are threefold. First, scholarly explanation of scaling requires an elaboration of the organizational mechanisms that generate outcomes consistent with our definition of scaling. This needs to go beyond mere discourses because important aspects of reality are not dependent on discourses as mechanisms (Bhaskar, 1975; Bunge, 2006). Second, a mechanism-based analysis of scaling requires a systematic treatment of counterfactuals. If scaling implies purposive action and its consequences, we need to explain the absence of unintended consequences of our actions, particularly those with negative outcomes. And thirdly, we need to develop an operational model of reality as an analytical tool that enables a more systematic exploration of social phenomena of scaling. Identifying relevant explanatory counterfactuals that are often unobservable requires guidance in the form of an analytical flashlight to avoid getting lost in a nebula of actual and potential idiosyncratic or irrelevant factors.

How Critical Realism informs this study

Critical realism as a philosophy of science was developed and substantially commented on by a number of scholars including Harré and Madden (1975), Bhaskar (1975), Collier (1994), Archer (1995), Sayer (1992), Lawson (1997) and others. Realist perspectives were also proposed as fruitful avenues for investigation in organization and management studies (Ackroyd and Fleetwood, 2000; Durand and Vaara, 2010; Fairclough, 2005; Johnson and Duberley, 2000; Reed, 2005; Tsang and Kwan, 1999; Tsoukas, 1989; Van de Ven, 2007; Whitley, 1984). A hallmark of critical realism is the commitment to a differentiated reality. Critical realists view ontology as constituted by an empirical, an actual and a real domain. The *empirical domain* refers to empirically observable events. These events constitute the outcomes generated by the operation of mechanisms. However, this domain does not exhaust reality. Critical realism differentiates two further levels understood as “deeper” levels of causality in the sense that in

these deeper levels causal explanation eventually “bottoms out”. The actually operating mechanisms and the empirical events they generate constitute the second dimension of the critical realist ontology, the *actual domain*. This domain is deeper from an explanatory perspective because causality flows from mechanisms to outcomes. The third dimension, the *real dimension*, aligns the whole causal chain from objects to outcomes. Mechanisms are enabled in concrete objects due to their internal structural configurations and their relations with other objects. In simple terms, the mechanism of running is enabled in humans due to the structural characteristic of having legs. That does not mean that everyone who has legs runs but that running is amongst the set of enabled mechanisms of that human individual. The mechanism of flying is not enabled in human beings and thus the total set of mechanisms is finite although we do not know its extent ex-ante observing empirical manifestations. Enabled mechanisms also reflect social structures. For example the hierarchical social structure of manager and direct report engenders in the manager an enabled mechanism to fire the direct report but not the other way around. The real domain thus represents a state space that is constituted by the specific set of enabled mechanisms, the specific subset of actualized mechanisms and the set of outcomes generated at a specific time and place.

A commitment to this differentiated ontology necessitates the view that – outside of strictly controlled experimental systems, i.e. in the real world – there are multiple possible correlations between a cause and an effect. The reason is that the three ontological layers are irreducible to each other and are out of phase with one another (Lawson, 1995). Observable outcomes are out of phase with mechanisms because many mechanisms operate in parallel and may reinforce each other in unexpected ways or cancel each other out. Mechanisms may operate below a critical threshold and thus their effects remain unrealized. Mechanisms are also out of phase with the structures that enable them. They may not be adequately triggered or social actors may choose to act in different ways in the same situational context. The consequence is that regular realizations of causes and effects, i.e. event regularities, are not expected outside strictly controlled experimental systems. The best we can expect are what Lawson (1995) calls “demi-regs”: temporary event regularities that create short-lived and spatially restricted patterns: doing A may sometimes create outcome B or doing A has a tendency in certain situations to create outcome B.

Critical realism highlights the indeterminacy of real world open systems and thus poses a challenge to our definition of scaling as a form of event regularities. But the differentiated ontology also opens the door to an analytical operationalization of the views of sociologists about “unintended consequences” (Merton, 1968), “derailing factors” (Portes, 2000) and counterfactuals (Tilly, 1996).

METHODS AND DEVELOPMENT OF AN ANALYTICAL MODEL

Realist mechanism-based explanation begins with a clear specification of an explanandum (Mayntz, 2004). Our starting ground is not a concrete observation but rather the definition of the type of organizational phenomena we wish to explain – in our case it is a particular definition of scaling as a robust set of event regularities. We are conscious that our definition of scaling may

not be valid for other definitions or uses of the concept of scaling. Following a critical realist research tradition our principal mode of inference is retrodution which postulates and identifies mechanisms that are capable of producing or actually produce certain events (Sayer, 1992). Retrodution takes us *“behind the surface phenomenon to its causes, or more generally from phenomena lying at one level to causes often lying at a different deeper one”* (Lawson, 2003, p. 80). Realist explanation requires that *“Science is the illumination and elaboration of the structures and mechanisms that govern the events of experience”* (Lawson, 1995, p.13). This means that we need to specify the particular characteristics of an organization that would explain how scaling can occur. To enable a more transparent and systematic process of retrodution, we develop a generic analytical model as a tool for investigating the phenomenon of scaling. We build on the important perspective of using models as analytical instruments (Morrison and Morgan, 1999). However, our specification of an analytical model does not mediate between explicit theories and reality as it is used by Morgan and Morrison (1999). Rather, our analytical model mediates between a particular philosophy of science (critical realism) and reality. The model is thus our way of operationalizing the crucial critical realist mode of inquiry. In the words of Sayer (1992, p. 122): *“Realist philosophy reflects upon the conditions which must hold if regularities are actually to occur, that is, it asks what a system and its constituent objects must be like for regularities to be produced.”* For the purposes of our study, the analytical model thus specifies the characteristics of an organization that enables it to produce the event regularities implied in our definition of scaling.

We then scrutinize the model empirically (Leca and Naccache, 2006). We do so by basing our analysis on one in-depth case study, creatively using alternative data sources and identifying factual and counterfactual mechanisms *“by continuing to ask the question why and collecting more data until we believe we have an explanation”* (Easton, 2000, p. 217). The objective is to *“...render a phenomenon intelligible. Mechanism descriptions show how possibly, how plausibly, or how actually things works.”* (Machamer et al., 2000). One of the key challenges in mechanism based explanations is the fact that important mechanisms may not be observable (Hedström and Ylikoski, 2010; Bunge, 2006; Gerring, 2007). This may be true of actually operating mechanisms but even more so of counterfactuals that remain in a non-actualized state and therefore represent system potentials rather than observable objects. To overcome this difficulty Gerring (2007) proposes to narrow the scope of investigation and to get closer to the micro-foundations of mechanisms. We make this operational by searching for mechanisms at their actual level of instantiations. We directly capture the reflections of organizational actors on their work environment. In-vivo codes in the form of personal reflections on the links between actions and outcomes from multiple sources are crucial for identifying actual and plausible sets of mechanisms. We also observe actors in their work environment to identify habitual tasks that the actors may not reflect upon. Using our analytical model (see below and Figure 2) we explicitly code for sets of relevant actors, enabled mechanisms as properties of these actors or their relations, how mechanisms are triggered and how they generate their desired consequences. Furthermore, this allows us to postulate important counterfactuals. They are captured as those events that actors have observed as causes of unintended and undesired consequences in their

immediate work environment. We also code for the robust mechanisms by which actors disable counterfactuals, suppress the realization of counterfactuals or neutralize their consequences.

We go back and forth between sources of data evidencing robust outcome patterns and data indicating the mechanisms that generate the outcomes. That way, we identify a set of actual and plausible mechanisms that are explanatorily relevant. We then code-up sets of related mechanisms into smaller sets of key processes. Processes are sets of related micro-mechanisms that create robust higher-level organizational outcomes. Our interest is in identifying those processes that directly generate the set of explicitly desired outcomes in the form of organizational objectives. The processes are thus causally efficient in explaining outcome regularities. They render the phenomenon of organizational scaling as used in this paper intelligible. Figure 2 summarizes our coding approach.

As an illustrative example, we chose to study Aravind, one of the largest eye hospitals in the world that cater to poor patients as an extreme case to illustrate the usefulness of our model. Scholars have provided detailed descriptions of Aravind's ability to consistently produce tangible and clearly measurable outcomes over many years (De Véricout and Sousa Lobo, 2009; Rangan and Thulasiraj, 2007). Our analysis benefitted from superior quality of data due to the privileged access of the research team to the organization over an extended period of time (from October 2004 until November 2009).

When gathering and analyzing data, critical realist researchers use discourse analysis to get as close as possible to actual micro-mechanisms. Central to this approach is that researchers take a critical stance (Leca and Naccache, 2006) and observe actors' actions and practices beyond the discourses they develop (Archer, 2002). We have conducted multiple fieldtrips to India and triangulated data from more than 51 interviews by interviewing (1) the same person several times and in different (group or geographical) settings², (2) a number of informants on the same practices and events, and (3) experts, external collaborators and members of support organizations. We corroborated interview data with reports and studies on the focal organization conducted by independent researchers, direct observations of practices by four researchers at different times, participation of two researchers in management meetings and retreats of the management team, archival data including internal reports and presentations including performance data on the various hospitals, and physical artefacts. Appendix A provides a summary of interview data.

² For instance, we interviewed the Executive Director seven times during the course of our study. The interviews were conducted by different researchers, at different occasions and in different locations/continents.

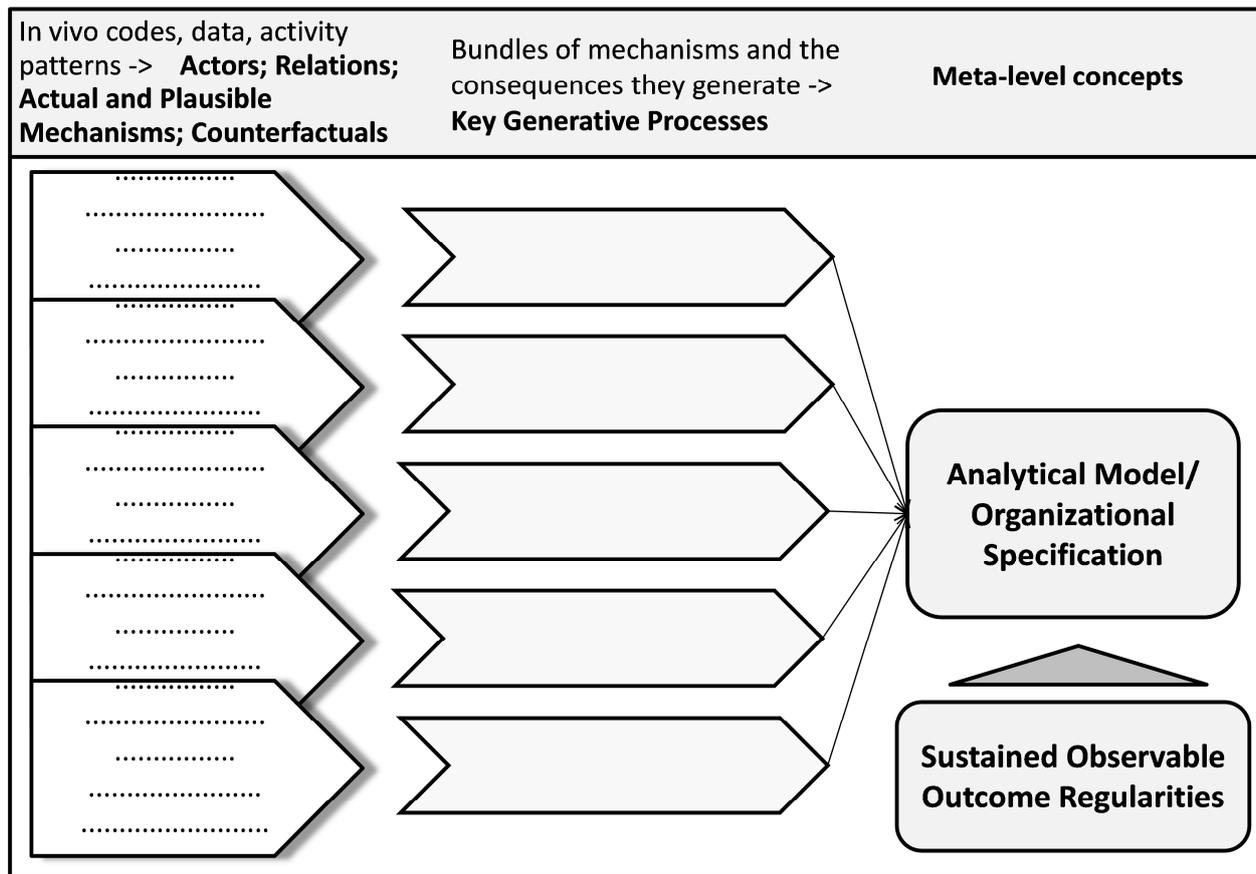


Figure 1. Coding Scheme – generative causality works from left to right (actors, their relations, enabled set of mechanisms, the ways they are triggered and the situative consequences they generate) while analysis works from right to left (observed regularities; specification of an analytical model; retroduction of causally effective mechanisms and processes).

Analytical Model

Building models as analytical instruments requires an explicit treatment of two dimensions of validity, epistemological validity and empirical validity (McKelvey, 2002; Seelos, 2010). We start building the model from the dimension of epistemological validity. Engaging a realist philosophy of science gives us a generic set of concepts and relationships. This enables us to create the basic architecture of a generative model, i.e. an analytical map of an organization that links actors, mechanisms and outcomes. The model receives additional specification by the analytical purpose of explaining event regularities. It thus specifies what an organization needs to be like for robust and sustained event regularities to occur. We engage the work of Bunge (2006) who defines the minimal required set of specifications for a material system (such as an organization) as: "constituents", "structure", "mechanisms" and "environments". To develop our analytical model we merge this set of system specifications with the critical realist ontology by making the following adjustments:

1. We limit "constituents" to human actors; all non-human constituents such as machinery are explanatorily unproblematic for the purpose of this study because they do not create systemic unpredictable variance that undermines the creation of event regularities.
2. "Structure", following Hodgson (2007) and Tsoukas (1989), refers to the set of human relations that have both enabling and constraining effects on the generation of mechanisms.
3. Because mechanisms are enabled by the structures of objects and their relations, the variables "actors" and "structure" together determine the system potential and the feasible set of mechanisms that can be generated (Harre and Madden, 1975; Bhaskar, 1975).
4. "Mechanisms" refer to the set of distinct activities that make a system "*what it is and the peculiar ways it changes*" (Bunge 2006, p.126). The mechanisms of interest are thus the causes of analytically relevant consequences. Consequences can themselves be causes, i.e. when mechanisms are a part of a causal sequence or process.
5. Because organizations are neither isolated nor independent of their environments, our model needs to specify the relevant set of internal and contextual actors and mechanisms that determine the outcomes an organization can achieve.
6. We thus treat the variable "environment" as the relevant actors, structures and mechanisms in the task environment of the focal organization.

While actors, structure, mechanisms, and environment specify the content of our framework, we need to integrate the specific coupling between these variables and outcomes. We map the above variables to the differentiated ontology of critical realism comprised of the real, actual and empirical domains as depicted in Figure 2. Our definition of scaling implies a robust set of event regularities between actors, mechanisms and desired outcomes. Therefore, our model needs to specify the organizational conditions for robust event regularities to be possible. We thus treat the empirical observation of consistent achievements of concrete organizational objectives as the explanandum, i.e. the event regularity, and actors, structures and mechanisms as the explanans.

A central aspect of our analytical framework is Harré and Madden's (1975) notion of enabled mechanisms. To the extent that desired mechanisms actually generate the organization's objectives they need to be enabled. 'Enabled' means that they are part of the set of properties that define an object, i.e. properties that reside in system-relevant actors and/or their relations. For example, doctors in an eye hospital need to have training in cataract surgery if providing this service is an organizational objective. For Aravind as an organization the desired mechanism of doing eye surgeries and generating the desired consequence of restoring sight is enabled through proper training of eye doctors or through hiring trained eye surgeons. However, enabled mechanisms constitute a potentiality that may not be realized (Bhaskar, 1975). In our example, the enabled mechanism of eye surgery also needs to be triggered, for example by the presence of patients, the availability of proper equipment or by an adequate incentive system that motivates the doctor actually to do surgeries on a regular basis. Furthermore, realization of the expected outcomes of mechanisms also requires that negatively interfering mechanisms, i.e. undesired mechanisms, are disabled or suppressed. Otherwise enabled desired mechanisms may not be

triggered, or if triggered may be neutralized or dominated by undesired mechanisms, and outcome regularities cannot be achieved on a sustained basis as indicated by the dotted arrows in Figure 2.

Bhaskar (1975) refers to the condition of any system where a tight coupling of actors, mechanisms and outcomes exists and persists as a closure condition. In the natural sciences “closure” is achieved by setting up controlled experiments so that a robust coupling between objects, mechanisms and outcomes can be observed. As Bhaskar (1975: p45) explains: *"The experimental scientist must perform two essential functions in an experiment. First, he must trigger the mechanism under study to ensure that it is active; and secondly, he must prevent any interference with the operation of the mechanism. These activities could be designated as 'experimental production' and 'experimental control'."* The term “closure” does not necessarily refer to a closed system. Closure may be achieved more easily in a closed system because blocking any interference from external factors limits the set of possible counterfactuals. However, since social phenomena such as organizations occur in relatively open systems (Bhaskar, 1975; Archer, 1995), the achievement of organizational closure requires managerial effort in the same way that *"...the experimentalist's task is to manipulate the entire experimental system, so as to manufacture the desired interrelationship between independent and dependent variable. The experimentalist is indeed a system builder and the crucial evidence is produced not by controlled observation but by work."* (Pawson and Tilley, 1997, p. 60). We thus conceptualize organizational closure as a required organizational characteristic for sustained outcome regularities. Our analytical model in Figure 2 is thus a model of organizational closure – a specification of the generic organizational conditions that must be assumed for scaling as defined in this study to be possible.

The analytical model helps us to look at the world in a systematic fashion. Its epistemological validity is defined by its grounding in a realist philosophy of science. Ontological validity of the model requires specification of the relationship between the constitutive parts of the model and real things as part of a focal phenomenon. Demetriou (2009) notes that our ability to refer to actual mechanisms as they occur in reality is limited and needs elaboration by providing rich “commonsense narrative”. Rich narrative of the focal phenomenon is particularly fruitful for investigating counterfactuals. They are an important part of our model that is not visible and usually not captured by empirical investigation. One important channel to discovering relevant counterfactuals is when actors refer to examples of unintended or surprise events and/or when they highlight ongoing efforts to prevent or suppress them. Our analytical model illustrates how empirical events are also "caused" by the absence of enabled mechanisms that if triggered and enacted could have prevented the realization of that event.

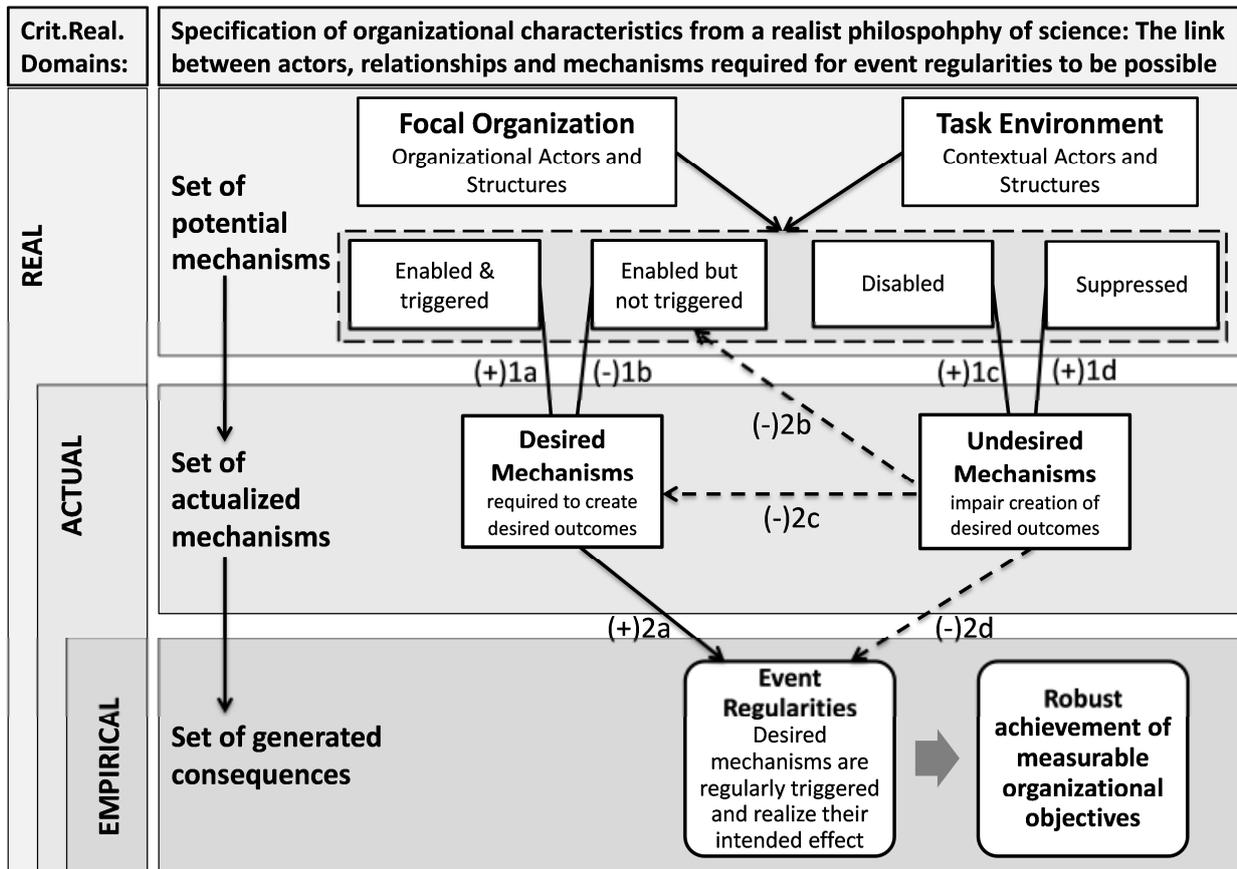


Figure 2. Analytical Model – illustrates the way we operationalize the three domains of reality of critical realism (real, actual and empirical) for organizational analysis. The model specifies how observable event regularities, implied in our definition of scaling, are caused by the action of desired mechanisms (positive effect 2a). They are also “caused” by the absence of potential undesired mechanisms that otherwise could negatively impact outcomes (negative effect 2d), the operation of desired mechanisms (negative effects 2c) or prevent that enabled desired mechanisms are triggered into action (negative effect 2b). The unique identity or constitution of actors as part of the focal organization and its task environment as well as their relations to each other determine the relevant set of enabled desired and undesired mechanisms. For a robust level of event regularities to be “produced” by the organization, desired mechanisms must be triggered on a sustained basis (positive effect 1a) and interference with this triggering (negative effect 1b) must be disabled (positive effect 1c) or suppressed (positive effect 1d).

ANALYSIS OF SCALING AT ARAVIND EYE HOSPITAL

We illustrate and complement our analytical model with a “rich narrative” as outlined in the methods section. The object of the narrative is Aravind, an organization that fits our definition of scaling very well. Iterating between the analytical framework and Aravind’s rich empirical

history enables us to identify a number of specific mechanisms underlying their scaling efforts and to generate more generic propositions on scaling mechanisms.

Aravind Eye Hospital

Aravind in India is the largest group of eye hospitals in the world. The group's mission is to eradicate unnecessary blindness and it has pioneered a novel approach to delivering eye surgery for cataracts that integrates free surgery for the poor as a major strategic objective. Cataracts affect millions of people all over the world and are routinely treated in wealthy countries by replacing the clouded natural lens of the eye with an artificial lens to restore sight. However, the market prices of lenses and surgery have traditionally been out of the reach of poor people. Partial or full blindness limits or prevents the ability of the poor to engage in economic activity. In addition to the personal hardships of the blind in poor countries, limited social welfare and insurance systems put a severe burden on the supporting families and their small private savings causing significant economic losses and social and individual suffering. According to the World Health Organization, the economic costs of blindness are estimated at US\$25 billion. Access to quality eye care and appropriate technologies would make up to 80% of blindness preventable or curable (World Health Organization, 1999; World Health Organization, 1997).

In 1976, Dr. Govindappa Venkataswamy ("Dr. V" as he is called at Aravind), a retired ophthalmologist, founded Aravind, an eleven-bed eye hospital, in the city of Madurai in the Indian state of Tamil Nadu. Over the years Aravind has expanded to a group of five hospitals that currently perform over 300,000 eye surgeries annually and provide eye care services to more than two million outpatients. While more than half of the high-quality eye surgeries are provided to the poorest for free, Aravind manages to generate significant operating profits (see Figure 3). Profits are used to invest in capacity building and increasing the scope and scale of Aravind's activities. This sustained growth of the delivery of a highly standardized service over two decades constitutes an event regularity that requires closure as an explanation from the perspective of critical realism.

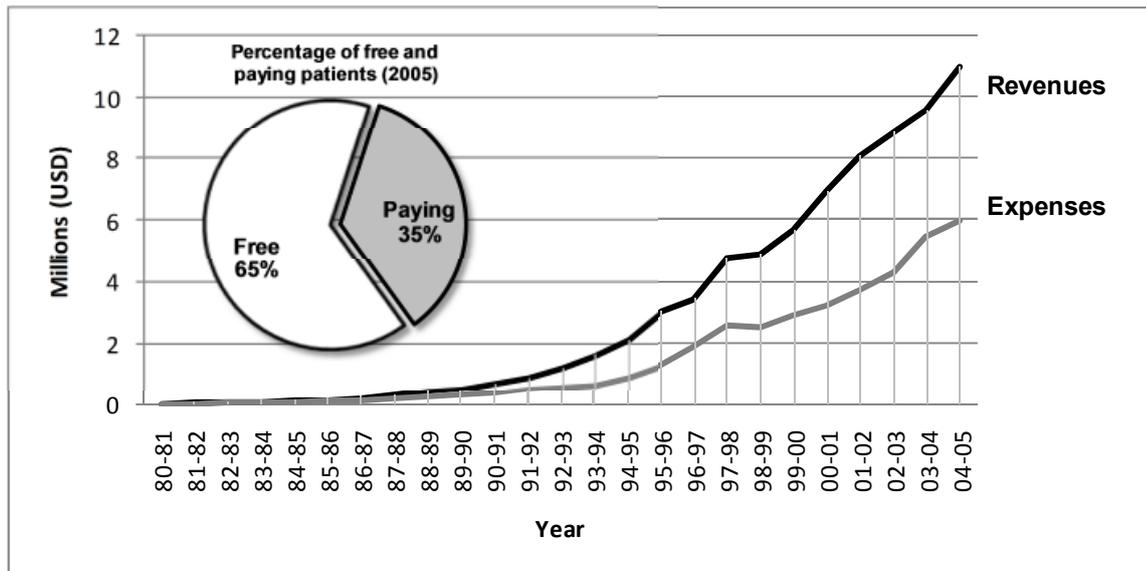


Figure 3. Revenues and expenses of Aravind between 1980 and 2005 and percentage of free versus paying patients treated in 2005. Note: Data provided by R. D. Thulasiraj, Executive Director of the Lions Aravind Institute of Community Ophthalmology.

How Aravind achieves scale

The mission of Aravind is to eradicate needless blindness. The core objectives defining Aravind's strategy are to provide large volumes of high quality cataract surgery; to achieve and maintain a high ratio (70:30) of treating poor non-paying patients to paying patients; and to achieve profitability. The strategic objectives constitute the outputs or direct consequences of their operation. Eradicating needless blindness is the outcome or consequence of generating these outputs over time. Scaling is thus an inherent requirement of this strategy. To achieve the outcome of eliminating needless blindness output needs to increase until a steady state can be achieved where all existing cases of cataract have been treated and the levels of output matches the levels of new patient cases.

Here, we highlight a number of mechanisms that generate the increasing scale of Aravind in achieving its mission and the ways in which these mechanisms are enabled and triggered. We have clustered the many individual mechanisms into five core processes. These processes explain how outputs are generated at an increasing scale. We thus label them generative processes. They are constituted by bundles of desired mechanisms and the ways how Aravind disables or suppresses potential undesired mechanisms. We also provide a summary table of the main counterfactuals we have identified (Table 1 in the annex).

Generative Process (1) – maintaining focus on a limited set of core services over sustained periods of time

From the beginning Aravind has focused on cataract treatment as its prime service. The focus on cataracts is relevant given the mission of eradicating needless blindness as an Aravind doctor confirms: *“Blindness is growing. About 330,000 every year is [the] incidence of cataracts in this area. [The number of] surgeries done in this area was about 75,000. Every year, a backlog of about 225,000 more blind people is pulling the society back”* (Doctor). However, the many needs in a context of large-scale poverty may tempt organizations to expand their scope and thus seek to "scale" their activities across many dimensions of needs. Examples such as BRAC in Bangladesh, Sekem in Egypt, Hand in Hand in India exemplify this tendency.

From the perspective of our analytical model, expansion of scope expands the set of organizational and external actors, relations and thus adds complexity. This may lower the levels of control over too wide a range of organizational and contextual variables, which, in turn, constitutes a loss of closure. Expansion in scope is thus an enabled counterfactual that might be triggered in several ways: the emotional pressure on members of a profitable social enterprise not to ignore many other important social needs in rural India; the curiosity of eye doctors to engage in other activities than cataract surgeries; fears of deskilling of eye doctors by just doing cataract surgeries (further discussed below). To suppress the tendency of scope expansion, Aravind explicitly maintains its original goal of eradicating needless blindness and cataract surgery. Before his death, Aravind's founder, Dr. V, encouraged a number of relatives, many of whom are eye doctors, to work within the organization. Family members occupy key positions throughout Aravind. Their strong bonds and close relations to each other and their status within the organization keep Dr. V's legacy in the form of Aravind's dedication to its mission alive. Another mechanism to suppress tendencies for "mission drift" is the recent decision and public commitment of Aravind's board to a new "stretch goal": to build the capacity to provide one million cataract surgeries annually by 2012. The commitment increases the pace and requirement of efficiency for cataract surgeries and leaves little opportunity for individuals to pursue other activities. Thus, two thirds of all eye surgeries performed by Aravind are cataract surgeries, while the other third combines another eleven different types of eye surgeries.

Generative Process (2) – providing robust levels of high volume treatments

Because of the high incidence of blindness, achieving high-volumes of cataract surgeries and growing its capacity is essential to the success of Aravind's mission. As summarized in table 1 there are a number of counterfactuals that challenge the realization of Aravind's mission. We have scanned the database of Aravind's advisory work with a number of other eye hospitals in several countries. The discrepancy with Aravind's performance levels is striking (data not shown). Their levels of surgeries are either much smaller in terms of surgeries per eye doctor (a productivity problem) or fluctuate significantly from year to year (loss of event regularities implying insufficient closure) or both. In contrast, Aravind is amongst the largest eye hospitals in the world and their ability to deliver high-volume services at increasing scale is extremely robust (see Figure 3).

Because our analysis is geared towards explaining closure, the resources of interest here are primarily human resources such as doctors, nurses and family members. It follows from our analytical framework that outcome regularities depend on the ability of Aravind employees and relevant external actors to have the particular mechanisms enabled that are required for the provision of high volumes of high quality services. Furthermore, since all actors have more than just these desired mechanisms enabled, Aravind needs to find ways to prevent the actualization of potential undesired mechanisms. We identified three main organizational mechanisms that Aravind deploys to achieve this: the building of reputation and trust in the task environment; the internalization of training that emphasizes skills and values; and establishing a rigorous selection process for various actors.

The reputation and trust that Aravind has built over the last 30 years has enabled the organization to ensure a continuous supply of nurses and doctors. Recruiting girls from rural India and training them as professional nurses required Aravind to overcome some limiting norms and traditions. For example, girls are supposed to get married at the age of fourteen and are not encouraged to leave their villages to work in cities such as Madurai. *“Also it is very difficult to recruit girls in this part: here, the psychology is that there are social arrangements [marriage] and all. Families were afraid to send their daughters. Now they’re comfortable with it. Yes, that was one of the advantages of Aravind in the south area: they knew the organization. If not to other organizations, they’d send to Aravind. [...] Now, we’re getting lots of applications”* (Hospital Administrator). It took Aravind several years to gain legitimacy as a trustworthy institution: *“The families of those girls, they feel that the girls are safe, they are serving for a good cause, they have values of [the] culture of Aravind, and they’re serving the people. So, they are very happy. And first year, there is a girl from the family, next year relatives are here, because it’s good here”* (Chairman).

While legitimacy and reputation ensure the supply of resources/actors, it is the content of the training, i.e., a combination of skill focused and values based training that generates the necessary attitudes to trigger the right desired mechanisms within Aravind's system. *“Training is bi-directional. Our paramedics and doctors have to be professional and very efficient and competent. [The next most important thing is having] ... ethics equal to the values of Aravind. How to perform beyond 100%. Hard work takes you to 90%, knowledge takes you to 96%. But attitude will take you to 100%”* (Chief Medical Officer).

A potential undesired mechanism therefore lies in the fluctuation of training personnel and inconsistent training programs. Observing Aravind’s extreme levels of efficiency and work-intensity in their hospitals revealed a situation that severely constrains key personnel to make time available for training. The strong focus on the mission and the many patients that are waiting at all times for treatment create a plausible counterfactual of doctors not being committed to training. Aravind disables such tendencies by the direct involvement of family members in both skills and values training. The following description underlines this process whereby family members become the carriers of skills and values and their involvement disables undesired

mechanisms: *“Basically first of all the training that I received at Aravind. ... I have observed many surgeries. I have observed different techniques, machines, all the aspects to run a hospital. ... The next thing I was lucky enough to meet Dr V. Always he emphasized on the attitude and knowledge adaptation. [...] Other organizations, I don’t know how they give emphasis on the vision and mission. Maybe they are on [...] paper, but in Aravind it’s not like that, it is not a piece of paper. It is actually taught, trained and they ask us to perform it”.* (Hospital Administrator)

As important as the quality of training to achieve outcome regularities is the decision of who is hired. How disciplined and aligned with the Aravind culture and processes will the individual actors – nurses, administrators or doctors – be in performing their tasks? In order to disable potential undesired mechanisms or avoid hiring the “wrong people” Aravind engages in a meticulous selection process: *“When we select a person, both consciously and unconsciously, the most important criteria is organizational fit. At every level. And the more senior they are, the more rigorous is the assessment process. Like, for example, taking a doctor, because a doctor’s position by default is an influential position in the organization: people listen to them, the nurses and the people. If they set wrong standards, that would kind of dilute the organization, so we put a lot of emphasis there. We also want those candidates to feel comfortable in this work environment. [...] in selecting senior people, doctors and all that, we would often have them spend about three or four days with us before we make a decision”.* (Executive Director).

A widely used mechanism to suppress behavior that is not aligned with Aravind’s important cultural elements – compassion, transparency and integrity – is deemphasizing the reliance on individuals: *“We trust the common man rather than a VIP. We do not go for titles and positions. I can say to everyone, “I studied in Harvard 40 years back”. Who cares; it is all about my behavior. You will never see our titles anywhere: for the patients, I am Dr Natchiar, that is all. Only when they look at me, they think I am a senior, I am an old lady”.* (Senior Founding Member). Reflecting on an instance where Aravind had to intervene to discipline a doctor, the executive director explains: *“Very, very rarely we’ve asked people to go. Maybe in the 30, 35 years, I can only think of maybe one or two instances. Very rare [...] Like even recently, I had to counsel an arrhythmia specialist, because he was not becoming a team player. He likes to publish. That is good. We appreciate that. And he is a good doctor. But then there was a time when he went overboard, like throughout the day he would want to do internet browsing. There are patients waiting. If internally they would tell something at the department head level, he would kind of do it but not with the spirit. Then I had to sit and chat with him, really making him understand the biggest loser is him, not us. Ultimately, it is you who is wasting time, and the biggest impact will be you, because this institution will go on after you leave.”*

Habit formation to ensure disciplined and uniform behavior also spans across all task domains as an important way to enable and trigger desired mechanisms without much conscious consideration, and to disable undesired mechanisms: *“Northern girls are very loud and expressive compared to the southern Indian women. But in the hospital setting, you cannot afford*

being very loud. Everybody knows that operating skill is a skill that if you repeatedly do it, you will get it. But what is important is not the skill, value and attitude, that makes the difference. [...] My first job is to make them de-learn what they have learned during their undergraduation. After de-learning then we inculcate the Aravind poison” (Doctor).

Generative Process (3) – provision of high-quality surgery to 70% of all patients for free
Aravind provides the same quality treatment to both paying and non-paying patients. This policy disables or suppresses a number of undesired mechanisms. Having only one type of surgical procedure for all patients eliminates ambiguity, misinterpretation or potential conflict for nurses or doctors about what type of surgical procedure would be adequate for any given paying or non-paying patient. Many nurses and doctors have joined Aravind in the belief of its mission of fairness and not treating poor people as second-class patients. Staying true to this mission thus suppresses potential feelings of distrust or cynicism among employees which would counter Aravind's need for high employee dedication and motivation: *“Every employee is very proud of us. Even a gardener [...]. When they feel pride, they feel without them it's not going to function. When there is this right [culture], I think so many things follow” (Doctor).* *“[If the] patient is satisfied with me, I am happy too. Without these things, we can't develop our hospital. Hard work is needed here, we [all] contribute” (Nurse).* The commitment to serving all patients regardless of income or background and systematically suppressing any tendencies to prioritize paying over non-paying patients (for example, to increase profits) are crucial mechanisms for Aravind: *“Why is Aravind unique? We have a huge patient load. We have so much technology. We have 250 doctors. We have all. But as I said, we are modest. We don't cheat our patients. We are not greedy for money. We always do the best for the patient. We always respect people” (Founding Member).*

The decision to keep the ratio of non-paying to paying patients at 70:30 is highly relevant given Aravind's mission. Cataract services in India's public hospitals are financially out of reach of the poor. However, being poor and blind is a dramatic burden in developing countries because of the lack of welfare or insurance systems. Therefore, the need for cataract treatment in the poorest part of the population is very high. But the needs of the poor do not automatically constitute a real demand. For Aravind to generate this demand, a number of undesired mechanisms in Aravind's task environment need to be disabled or suppressed. Many poor people consider blindness an age related or otherwise normal event so the idea that this can be treated does not naturally occur to them. Many are also distrustful towards doctors and perceive a visit to a hospital as a disruption to their regular day to day lives. A further complication lies in the necessity of patient compliance, which normally includes repeated visits to a hospital for diagnosis, surgery and post-surgery monitoring. Lack of compliance is a critical undesired mechanism preventing the generation of expected outcomes. Aravind enables a sustained level of large-volume demand generation from poor patients by organizing eye camps in rural villages. Because success depends on suppressing distrust, Aravind partners with local community groups or local politicians that have the trust and respect of the rural poor. Aravind has built up a large network of community partners to generate a robust structure for enabling and triggering camps on a

routine basis, thus generating sufficient demand for its high-volume growth strategy. Non-compliance is disabled by taking patients through the whole process in the shortest possible timeframe: *“We have a system which would make sure that we would try to see off-station patients quickly. [...] If it is eye-camp, they really want to complete the loop in one visit, the whole service loop. If they want glasses, we give the glasses right away. If they need surgery, we have buses waiting and then taking them back, bringing them back, because there’s too much hassle to make people come again and again.”* (Executive Director). This is also why the commitment to high quality is so important: *“As a policy we don’t do any marketing, but what we do is, we do quality: our quality speaks. It’s the word-of-mouth that helps us. [...] When you have your surgery done, when you have your satisfied patients that is your marketing tool. They go to the community and spread their satisfaction. It’s not only the surgery it’s the quality of services we provide”* (Hospital Administrator).

A further undesired mechanism is the potential of doctors to earn extra money from treating more paying patients, thus neglecting treatment of non-paying patients to the detriment of the 70:30 ratio. This is a common practice in India’s hospitals. To disable this potential, Aravind functions on a strict no-appointment basis. Patients receive treatment from Aravind, the organization, not any particular doctor: *“In [...] India, the doctor starts practice, becomes a very famous one, and starts a hospital. He’s a single person; if he goes somewhere, the hospital is closed. There is no system if [the] hospital is based on one person only. What we are trying to do: people are going to the hospital, thinking that they’ll be treated the best. [...] Other hospitals are doctor based: if a patient came to see me, next time, he should come to me again. But in a system like Aravind, it’s not necessary. Today, I’m here, next time maybe not. Whoever is there will take care of him. If he wants to see a doctor, he has to come here in a particular date and wait for him. Because none of the Aravind doctors work [on an] appointment basis, we take whoever comes. This is the main difference from other hospitals: otherwise, some doctors may not get any patients while [an] other doctor is overloaded”* (Chief Medical Officer). In addition the constant presence of Aravind family members further suppresses any deviating tendencies, for example through fatigue from the daily routine or loss of motivation. *“Now what we are constantly telling them, that we should be sincere, that we should have passionate care, that 70% should be free and 30% paying and we should respect the patient. All those things you have to take them along and if they lose it everything is lost. They will not be able to run the institution successfully. Today Aravind is working not because of technology”* (Founding Member). Many at Aravind have noted that staying true to its mission is a crucial trigger that gets people up in the morning every day to perform: *“The organization is maintaining values through leaders playing the role-models. We become an example. For example I come here at 7 am in the morning, they come too even if there is no such rule”* (Chairman).

Generative Process (4) – achieving and maintaining high levels of operational efficiency

A key set of mechanisms drives operational productivity at Aravind. They include the dedicated commitment to standardization, the provision of real time performance measures, and incremental experimentation. Everything required to deliver high quality eye surgery is

standardized and coordinated. This involves the organization of eye camps for fast and efficient scanning of potential rural patients, optimization of logistics and transport of patients to the main hospital, the actual surgical procedures and the post surgical treatment (Rangan and Thulasiraj, 2007). Strict task specialization at every level of the organizational hierarchy enables steep learning curves and focused skill development. The pace of highly formalized processes triggers predetermined routine action. Time compression suppresses the ability to reflect on a situation and develop action alternatives which in an optimized system are likely to be dysfunctional: *"Every case is a replica of another case. The paramedical staff smoothes the work, the time lost between patients goes down. That is how the system runs. Paramedical staff, trained well, cut down the time. Performance level of the doctor is enhanced because his work is facilitated. A normal doctor would be able to do 6-8 surgeries an hour here. It's all about time management"*. (Chief Medical Officer). As a result, Aravind doctors are extremely productive. They provide roughly 10-times more surgeries per doctor per year than doctors in public hospitals.

Observing the strict task specialization resembling a Taylorist approach to human resource management generates concerns about potential undesired mechanisms such as alienation, boredom, rebellious behavior and high turnover at all levels. Aravind generates various organizational mechanisms to disable and suppress such tendencies. Nurses for example go through a uniformly structured two-year training program. Consequently, they have the same set of skills and can be replaced with minimal task interruption, which suppresses their incentives to rebel and disables the negative organizational impact of any nurse leaving. By providing uniform in-house training coupled with equal pay across task domains, Aravind is able to suppress sentiments of task inequality which could negatively affect commitment and work attitude and therefore compromise regularities in outcomes.

The extreme task specialization of surgeons creates concerns over deskilling and loss of marketability amongst doctors. This could trigger the undesired mechanism of highly trained eye doctors leaving for better paid jobs. In India, eye doctors are still bottleneck resources and are high in demand. Aravind's structural response to suppress such a detrimental effect involves managing several channels for engagement with the international community of leading ophthalmologists. Aravind invests in facilities that enable cutting edge research in a number of eye disciplines. And the large number of surgeries at Aravind provides doctors with unique access to special or rare cases: *"... we also need to have this external interface. Otherwise, we get too inbred. [...] people take part in lots of conferences [...] Then we also have the residents from the best universities in the U.S. come as part of their rotation. All these processes, almost through osmosis, kind of brings in best practice from outside into the system."* (Executive Director). And Aravind invests in their own training institute for ophthalmologists to neutralize levels of doctor turnovers that they cannot prevent.

A second enabling and triggering mechanism to enhance operational productivity is represented by the provision of stretch goals and the transparent and real time provision and exposure of task performance: *"Sometimes we even set some kind of targets, especially on quality, like pressure-*

reduced complication rates, or we should achieve outcomes of this standard. Then we have a fairly robust system for collecting information, analyzing, then reflecting on it. This is, I would say, a formal process, but which continually evolves as well, which you can reckon is more internally focused." (Executive Director). In this case the process leading to outcome regularities spans across hierarchical levels, i.e., performance is reported and made public across task domains and hierarchies: *"We do a few things during the course of the year to kind of connect back to the founding values and principles of the organization. We have like competitions across the organization to kind of say how their work connects up with the mission of the organization, how this driving, or maintaining cars, or cleaning a toilet: how does it add up to reducing needless blindness"* (Executive Director). As this quote illustrates, task performance is directly and explicitly related to the mission and organizational objectives, which allows Aravind to disable and suppress mission drift and trigger continuous commitment to task performance. Furthermore Aravind disables or suppresses possible undesired mechanisms typically associated with competitive approaches to task performance. For example, during our field visits we observed on several occasions that numbers of surgeries performed by each doctor are put on a wall at the end of each day. But the name of the doctor is not openly revealed, it is (semi) private information to inform the doctor about his performance in relation to how others are performing. Aravind explicitly avoids individualizing performance by for example, selecting a 'surgeon of the month' or 'nurse of the day'. This approach to reveal task performance suppresses a potentially undesired winner/loser atmosphere without compromising the motivational effect of revealing performance levels: *"Discipline is, I would not say it is a cultural value, but it is a process which is extremely important in a hospital setting, because it is a teamwork. For discipline, in terms of coming on time, or all the work. If those things are violated, then the whole system suffers. We kind of give a lot of weight to that."* (Executive Director).

A third mechanism to enhance operational productivity is incremental experimentation to optimize all desired mechanisms. But experimentation is tightly controlled and monitored to avoid potential undesired mechanisms, e.g. allocating scarce resources to doctors' pet projects. Experimentation is further discussed in the section "implications for scaling" below.

Generative Process (5) – achieving profitability

Profitability allows Aravind to be independent from fluctuating levels of donations, an important counterfactual that might be difficult to control with concomitant loss of organizational closure. Furthermore, profitability is an essential means to realize Aravind's mission. Apart from high productivity, generating demand from paying patients is thus a key desired mechanism. Building capacity to accommodate paying patients, for example by establishing hospitals with state of the art design and technology as well as offering private accommodation options, enables demand generation. Yet, why would paying customers choose Aravind over any other private for-profit hospital? Aravind eliminates undesired trends such as losing patients to competitors by offering slightly lower rates for paid surgeries, better food and superior post surgical service for paying patients. In addition to nurturing Aravind's reputation for high quality surgery, the authenticity

generated by sustaining an ability to treat 70% of its patients for free, has created a strong reputation for being a particularly caring organization that paying patients value: *"Last year, we did about 5,000 camp surgeries. Among that, 2500 [were] totally free of cost. Among that 2500, about 1000 [were] self-sponsored (by our group). Sometimes, a paying patient would like to sponsor 4-5 patients. When they visit the hospital, we tell them that some part of their payment goes to cover free patients, they ask how much is the cost for a free patient, we say 700, they say ok, I want to pay for 5-10 free patients"* (Director). Because of Aravind's commitment to high volumes and high quality, its doctors are running down learning curves very fast and Aravind has a pool of the best eye surgeons of the world.

Table 1 (Annex B) summarizes the main counterfactuals, their potential consequences, and the ways in which Aravind disables or suppresses them.

IMPLICATIONS FOR SCALING

We base the development of propositions on our analytical model and on Aravind data. The propositions are contingent on a context of large-scale poverty as is the case in rural India. The way we characterize that context has implications for our perspective of scaling. In the context of large-scale poverty, resources of many kinds tend to be scarce. Any non-productive use of resources has therefore high social opportunity costs. The uncertainty of innovation to create new organizational models may therefore not always be the most efficient and productive use of available resources in such a context (Seelos and Mair, 2007). This implies that efforts to make existing proven models better may be preferred over building new ones. Large-scale poverty as conceptualized here implies an opportunity for organizations to be extremely efficient and to deliver highly standardized products and services on a mass scale for sustained periods and without much adaptation or the necessity for radical innovation.

Scaling as increased productivity

The perspective of organizational closure has a number of implications for scaling. As discussed, from a realist philosophy closure is a requirement for scaling where scaling means the increased achievement of concrete and ex-ante specified organizational outcomes. The link is specified by the available knowledge or recipe of the adequate set of desired mechanisms known to create organizational objectives as their effect. But that knowledge or recipe is not sufficient in itself because it only specifies what "could be" and not what "will be". The latter is contingent on a number of undesired mechanisms that could prevent the results of operating desired mechanisms to be realized. Organizational closure implies knowledge about the required set of desired mechanisms and the ways in which possible undesired mechanisms can be disabled and suppressed. The analytical framework in Figure 2 enables to distinguish the concept of "degrees of closure". Higher degrees of closure thus correspond to better competencies of an organization to enable and trigger desired mechanisms in such a manner that their effects are regularly

realized. Higher degrees of closure can also be achieved by recognizing potential undesired mechanisms and learning how best to disable and suppress them. The many ways in which Aravind achieves closure in the section above highlight this important link between organizational closure and scaling. We therefore put forward the concept of "organizational closure competencies" as an important competence of organizations striving to scale. We derive the following definitional proposition.

Proposition 1: Higher degrees of organizational closure are achieved by enabling and triggering more desired mechanisms and disabling and suppressing more undesired mechanisms and thus generate higher levels of achievement of organizational objectives.

Knowledge about cause and effect mechanisms is a prerequisite for closure. Closure implies specification of the desired mechanisms required to achieve desired outcomes. We argue that the relation between knowledge and closure is an iterative process where better specifications of mechanisms enable higher degrees of closure. However, higher degrees of closure better resemble conditions that enable controlled experimentation. Controlled experimentation enables better isolation of the effects of newly introduced mechanisms and thus facilitates systematic learning. This lowers uncertainty about cause and effect relations, and strengthens the quality of the knowledge base or recipe. Performance thus improves by replacing desired mechanisms with substitutes that are more productive and by identifying and eliminating functionally neutral mechanisms or organizational slack. The latter is crucially important in environments of low munificence where many types of resources are scarce and their opportunity cost is high. Inclusive growth in such environments makes increasing productivity an essential scaling mechanism. *"The operating room staff meet every week. They would review the issues that they had. It could be sometimes some supply was not there when it was required, or suddenly some equipment failed, or there was a complication, or the patients sat way too long in the operating room. It could be any one of those things. Then, obviously, solutions to these problems are best with the people who are dealing with it. When they raise a solution and then you act on it, it makes things better. That is how we built an ownership for the process, and for the organization. When you do it on an ongoing basis, dealing with smaller and smaller problems, then you kind of move towards perfection."* (Executive Director).

Proposition 2: Closure increases productivity levels through a positive feedback mechanism between the quality of knowledge and higher degrees of closure.

In subsidized business models such as Aravind, where paying customers subsidize the service of non-paying customers, productivity increases have particular effects on performance levels. Given the constraint of financial self-sufficiency as an organizational objective, increasing the ability to serve poor "customers" that cannot pay market prices is determined by two factors: 1)

increasing the numbers of paying customers to generate excess profits to subsidize poor customers; and/or 2) increasing productivity levels. A simple calculation demonstrates the effect (see Figure 4). We use the Aravind model and assume that paying patients pay five units for a cataract surgery. Initially, productivity is low (productivity level 1 in Figure 4). Increases in productivity create lower marginal costs per surgery from an initial level of five units (with zero profits available to subsidize free surgeries). At the higher productivity level of 5 in Figure 4, the cost has decreased to one unit per surgery which enables four free surgeries for every paid surgery. A linear decrease in costs in this subsidized model generates non-linear capacity growth to service non-paying customers. This relationship is shown in Figure 4 below.

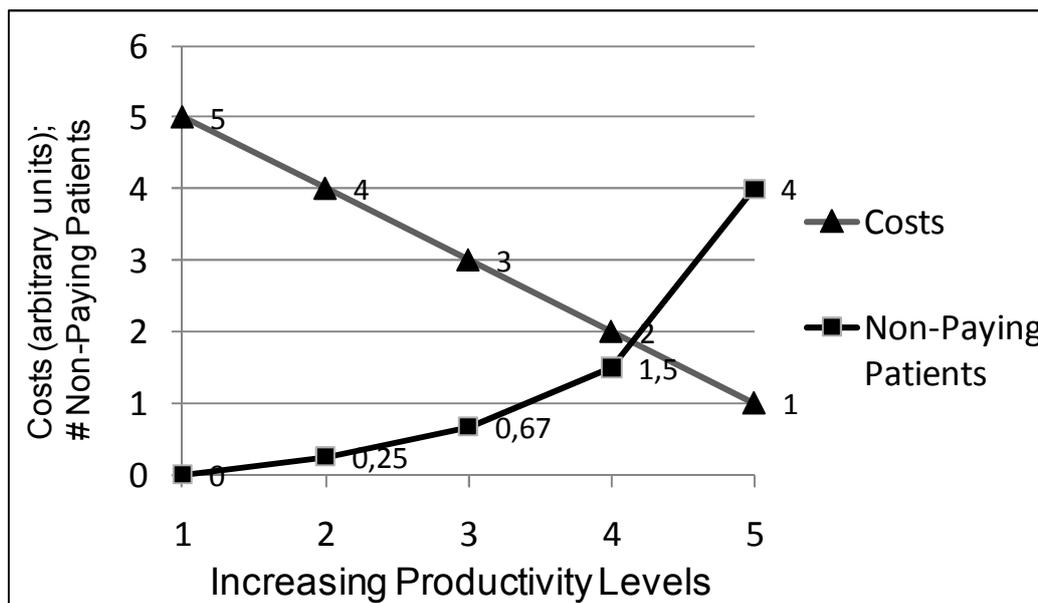


Figure 4. Relationship between productivity levels (defined as costs per treatment) and ability to fund treatments of non-paying patients per paying patient (who pay 5 cost units for one surgery).

For example, in the early 1990s Aravind decided to eliminate a dominant cost bottleneck – the highly priced intra ocular lenses required for cataract surgery. It founded Aurolab and thus brought the lens prices down from about 200 USD to about 4 USD. This greatly facilitated the ability to treat up to 50% of patients for free and still create the surplus required to fuel Aravind's inclusive growth model. *“Everybody thought when you do free work you will not become viable financially and Dr V [Aravind's founder] said, when you know your expenditure a little bit, when your productivity is better, when you have a good management system, you can be very viable. That’s our model. So the model became high volume, high quality, affordable cost. Volume was*

high so productivity became more and productivity became more and money came. When money came we could treat more people." (Director)

Proposition 3: For profitable business models that also serve the poor despite lower marginal revenues, a marginal increase in productivity has positive scale effects on the numbers of poor customers that can be served.

Scaling as increased size

An obvious scaling mechanism is the addition of resources to an organization. Closure allows an organization to identify resources that enable desired mechanisms and the organizational design enables the triggering of these mechanisms and ensures the realization of outcomes by disabling undesired mechanisms. Adding resources that enable desired mechanisms thus directly scales operations and achieves higher levels of outcomes. However, adding resources increases organizational size and this generates higher complexities that make it more difficult to maintain closure conditions. *"Lots of things are changing. As the system is getting bigger, you shouldn't get diluted. Somehow, as things are managed here, it's not diluted because of the close knit of the senior group that's always there and that binds you together and that makes you do things in a certain way."* (Director).

Proposition 4: Adding resources that enable desired mechanisms to an organization that has achieved closure has a positive effect on performance up to the point where marginal loss of closure due to organizational complexity equals marginal gain from desired mechanisms.

Scaling as replication

Proposition 4 implies replication as a mechanism to overcome the liabilities of scaling organizational size. Replication allows the establishment of a copy of the original model. This creates an opportunity to replicate at an organizational scale that still has positive marginal performance when adding resources. However, in environments of low munificence, resource constraints of all kinds are a challenge to replication. Thus, lower productivity levels make it imperative to first create the knowledge that enables higher closure and therefore increased productivity, before replication can be expected to make a positive overall contribution to scaling.

Reflecting on the need to replicate the whole system of Aravind not just modules such as technology, a founding member commented: *"Aravind is very particular. [...] our success is technology, team work, commitment, hard work, willingness to change; and we have our own values and culture. So that also should calculate into the human resource area; and whatever we focus [on ... for example] the attitude of the person, everything should be similar otherwise there*

is no use of the best technology. Everybody can go and operate, everybody can do better surgery than us, have a better building than us. The model of high volume, high quality, affordable cost, with compassionate care to all, that is our own model, which we need to transfer". Furthermore, Proposition 2 implies that higher levels of closure facilitate the replacement of desired mechanisms by better substitutes. This means that replicated organizations have an opportunity to maintain closure across the branches and thus diffuse new mechanisms rapidly and in the same manner. However, the ability to achieve fast dissemination of new mechanisms also requires higher degrees of closure as there will be less local interpretation of the feasibility or necessity for change and the new mechanism is more likely to work as expected. Aravind has four branch hospitals which regularly exchange comprehensive reports on finance, surgery performance and quality. A director at Aravind reflected on this sharing mechanism: *"All the finance gets updated here in Madurai [...] almost every day. [...] For example, there was a remodeling done in the glaucoma clinic in Coimbatore [branch]. They reorganized the patient flow. They were able to bring down the unnecessary waiting time from 2 hours to 18 minutes. That was very dramatic. So, the design of the area and the process changed. And that spread like wood fire across the system. That sharing happens. It was shared in one of the clinical heads meetings."*

Proposition 5: Given resource constraints, higher levels of productivity are positively associated with the organizational ability to achieve scale through replication.

Scaling as knowledge transfer

When organizations do not have the resources or do not want to replicate, knowledge transfer to other organizations is a potential mechanism for scaling. Our perspective implies that knowledge transfer cannot be expected to improve the performance of the receiving organization unless it has the closure competencies necessary to create desired outcomes as a result of integrating knowledge resources. An Aravind consultant reflected on the differences between managed care as a much more hands-on mode of knowledge transfer over longer periods of time and the work of Aravind consulting that focuses more on short-term engagements and exchange of best practices: *"In managed care, we can take some credit, when we are running it on a regular basis. But in consulting, we only act as a catalyst, we can't directly say 'it is because of me'. Because, once we act as a catalyst and they receive the ideas, they'll have to start implementing it from their side. But there are hospitals which have not done well because of lack of or frequent change of leadership or lack of availability of doctors or they're not following the processes they should"* (Director). Even the much more hands-on managed care approach to knowledge transfer is challenged by the inability to replicate the organizational circumstances required for the knowledge or recipe to realize its expected outcomes: *"In managed care, it's been a challenge of putting in the right structure, but we're able to put in the systems relatively well. But putting in the culture has been a challenge. I don't think we've succeeded in doing that"* (Executive Director).

Proposition 6: Knowledge transfer from an organization that has achieved closure will not enable similar levels of performance in the receiving organization if it cannot establish the same degree of closure as the original organization.

DISCUSSION AND CONCLUSIONS

The academic literature on scaling is in an embryonic stage. A recent review³ lists fifteen key references. Of these only three are published in scholarly journals highlighting the interest and need of practitioners for a better understanding of the concept of scaling. We contribute to the literature by providing a scholarly treatment of scaling. First, we focus on an explicit definition of the concept. Scaling in this paper refers to event regularities in the sense that doing more of A or doing A better creates more of an expected outcome B. Second, we apply the perspective of critical realism postulating that event regularities require closure. Closure is traditionally assumed to be restricted to artificial experimental situations. Critical realist scholars are pessimistic for the ability to generate closure conditions in social systems as is implied in our definition of scaling. Bhaskar (2008) states *"The social world is characterized by open systems where causal laws are out of phase with patterns of events and experiences; a constant conjunction is no more a necessary than a sufficient condition for a causal law."* Archer (1998) is concerned that *"Even in isolated environments, the nature of humans means that "closure" cannot be achieved"*. Thus the literature is ambiguous about whether and how it is possible to create organizational closure. While Tsoukas (1989) postulates the impossibility of constructing closure conditions in the social sciences, the same author also considers the opportunity to escape the indeterminacy of open systems: *"In other words, management must create conditions of organisational quasi-closure so that certain activities of interest are controlled [...] and particular results are obtained. Thus although the causal powers of management operate in open systems it is only when quasi-closed systems are constructed that a set of desirable regularities accrues."* (Tsoukas, 2000. p. 40). We therefore investigated the possibility of quasi-closure as an organizational characteristic required for generating sustained event regularities. Third, we identified an organization that has provided a highly standardized medical service over more than two decades as an extreme case of sustained event regularities. Reflection on original case data and theoretical literature enabled us to develop our central analytical model of organizational closure. The model facilitated a retroductive logic explaining how actors, structures, mechanisms and outcomes are robustly coupled to achieve defined organizational objectives. Our perspective also highlights the need for a counterfactual explanatory logic. This is operationalized by the organizational requirement for disabling or suppressing potential undesired mechanisms. This matters hypothetically insofar as their realization would constitute undesired mechanisms that would prevent the actualization of the effects of desired mechanisms and thus the achievement of defined objectives. Therefore,

³ Scaling Social Impact Research Project (2006) CASE, Fuqua School of Business, Duke University.

even if undesired mechanisms are not actualized and thus would escape empirical investigation, an explanation of organizational closure requires consideration of how they are disabled or suppressed. We propose that organizational closure is an organizational competence required for our definition of scaling. Closure competence refers to a distinct set of organizational knowledge about: a) the required set of desired mechanisms to achieve objectives; b) how actors and structures need to be selected and configured so that desired mechanisms are enabled and triggered on a sustained basis; c) the expected set of undesired mechanisms that actors and structures could generate in parallel or as an alternative to desired mechanisms; and d) how to disable or suppress undesired mechanisms so that the effects of desired mechanisms can be repeatedly and reliably actualized. Finally, we developed propositions about scaling that followed from our analytical framework and reflections on the illustrative case study.

Implications for practitioners

Our propositions on scaling mechanisms highlight the importance of the perspective of organizational closure in achieving higher levels of scale and productivity. Given the scarcity of resources in the context of poverty, productivity measures, i.e. outcomes generated from a given resource, are essential metrics for project evaluation. Funders of pro-poor growth models typically evaluate projects by isolated input or output measures. Our perspective requires an explicit evaluation of the coupling between input and output variables. Our model of organizational closure provides practitioners with a systematic framework for thinking about scaling and productivity. It also facilitates a more differentiated perspective on the types of resources required for scaling. Often, monetary resources, e.g. development funds, grants or donations are considered crucial factors for progress. However, our framework helps to identify critical resources based on identification of the desired mechanisms they are able to generate. Money often cannot buy these resources in the context of poverty and thus long-term efforts to provide bottleneck resources need to receive explicit attention. Aravind made a key investment in its own training program for productive eye doctors, a real bottleneck resource given its pro-poor growth model. This has implications for rethinking funding strategies and to make more realistic evaluations of development projects (Pawson and Tilley, 1997).

Our scaling propositions allow us to illuminate the link between resource availability and organizational closure. This link gives rise to several scaling options that may support decision making for managers as well as funders. When resource levels and degrees of organizational closure are low, the best option is to enhance productivity by learning how to increase closure. This requires a combination of focused knowledge transfer and small-scale focused experimentation.

Where both degrees of closure and resource levels are high, adding key resources based on the existing understanding of which resources drive desired mechanisms and desired outcomes is the prime scaling option. If organizational size due to adding more resources reaches a point where complexity causes a loss of closure the best option is to invest in replication of the model at another location. Our analytical framework supports replication decisions by enabling a

sophisticated understanding of contingency issues, i.e. to what extent actors and structures in a new location enable desired mechanisms and/or may undermine project success by generating too many undesired mechanisms. Thus replication is a tricky issue and works best in similar contexts.

Where degrees of closure are high but levels of resources are low investments in building key bottleneck resources are optimal to improve capacity of the original organization. In addition, knowledge transfer to another organization particularly in a more resource abundant location is a valid scaling option. Informed by proposition 6, we need to pay attention to the potential that adding knowledge resources to an organization that has lower degrees of closure may not generate desired outcomes.

When degrees of closure are low, as when organizations have not found solid working recipes and when levels of resources are high, the best option is to invest in further innovation and prototyping. Higher levels of resources are compatible with the uncertainty and failures implicit in innovation. Emphasis should be on critical evaluation of what works and what does not in order to limit allocation of resources to inefficient and ineffective recipes. Figure 5 summarizes these implications.

Degrees of Closure		Availability of Resources
Low	High	
<ul style="list-style-type: none"> • Experimentation to increase productivity 	<ul style="list-style-type: none"> • Knowledge Transfer • Development of key resources 	Low
<ul style="list-style-type: none"> • Innovation and prototyping to create better recipes 	<ul style="list-style-type: none"> • Adding resources to increase capacity • Replication 	High

Figure 5. Scaling options contingent on levels of resources and degrees of organizational closure

Implications for research

Despite its promise, critical realism has received limited explicit attention in research on strategy and organizations. Tsoukas (1989) has argued for the epistemological validity of idiographic case research from a critical realist tradition. With this study we hope to demonstrate the usefulness of this perspective for investigating the concept of scaling and to inspire others to expand and sharpen our analytical framework and the ways it can be used for theorizing.

We provide a contrasting perspective on scaling based on a narrow definition grounded in the perspective of closure. Others look at scaling from an open systems perspective that includes multiple interactions with a variety of stakeholders which expands the scale and scope of activities. It also includes evaluating outcomes much more broadly by focusing on indirect impacts such as spill-over effects as well as unintended consequences and different dimensions of value creation (Uvin, 2000). Therefore, our definition of scaling limits the generality of our findings to one particular meaning of scaling and to a unique context of large-scale poverty. Much more systematic and comparative research is required to improve our understanding of the contingent nature of the interplay between desired and undesired mechanisms in their ability to create empirical outcomes.

Finally, we wish to share our experience applying critical realist perspectives as stimulating very rich discussions amongst the research team and encouraging deep thinking which we found quite challenging and enjoyable. It made us reflect deeply on the usefulness and limitations of mainstream research methods and our own commitments to ontology and epistemology – an ongoing inquiry into our roles as researchers.

Appendices

Appendix A: Description of the Interview Data Collected

Number of hospitals visited	6 (3 branches: Madurai, Pondicherry, Coimbatore; 3 affiliated units: Lucknow, Amethi, Kolkatta)
Number of interviews conducted	51 (14 of the interviews involved teams or groups, e.g., the management team and groups of nurses)
Number of interviewees involved	63
Organizational roles of the informants	<ul style="list-style-type: none">• Chairman of the organization• Founding members• Executive director• Directors and managers of sub-units• Finance, infrastructure, training managers• Hospital administrators• Medical officers• Doctors• Nurses• Trustee representatives, external collaborators, members of supporting institutions
Duration of the interviews	10 min – 1 hour 40 min
Number of pages of interview data	More than 350 pages

Appendix B: Table 1. Identified counterfactuals, their potential undesired consequences, and Aravind’s mechanisms to disable or suppress them.

Counterfactuals – potential undesired mechanisms	Potential undesired consequences	Disabling/suppressing mechanisms by Aravind
Mission drift – satisfying a variety of important needs of the poor in the context of rural India in addition to cataract surgeries	Complexity, loss of closure, inadequate development of key resources	Doctors in key positions are relatives of the founder – keep the legacy alive as a group; Explicit stretch goal that uses an output metric requires focused capacity building
Lower quality surgeries for non-paying patients to save costs or as a consequence of framing them as “poor” and thus less demanding	Loss of positive reputation and thus loss of poor patient demand; Loss of reputation as a caring institution and thus lower paying patient demand; Loss of efficiency due to ambiguity what type of procedure for which patient to use; Cynicism and loss of commitment amongst employees over treating poor as lower-class patients	Explicit policy of equal patient treatment; Using the same doctors for surgeries; Doctors do not decide whether they treat paying or non-paying patients; Strictly hiring only caring doctors; Intensive training of their own nurses and doctors (technical and values); Doctors from the family in key positions act as safeguards of the Aravind values
Inability to acquire key resources or loss of key resources in a context of general resource scarcity: Tendency of girls to get married at age of 14 rather than working; Lack of eye doctors in India; Aravind trained skilled doctors leave for higher-paid jobs;	Inability to grow output due to low levels of doctors or nurses; Inability to keep a constant pace which generates inefficiencies and service disruption; Inadequate skills generate low productivity and higher error rates	Treating nurses well so that families want their daughters to work at Aravind; Establishing a dedicated training institute for eye doctors; Maintain relations with a large number of community partners that generate a constant flow of rural patients to Aravind; Establish their own factory for high-quality intra-ocular lenses;
Dependency on external funds to subsidize non-paying patients, e.g. developing fund raising strategy	Funders changing their capital allocations or intervening with Aravind’s model; Facing temporary situations of too little funds that slow down growth or too many funds that cannot be absorbed or could create organizational slack	Having an explicit profit target; Independence from individual donors; Providing high-quality surgeries and a caring reputation attract paying patients
Variance in behavior reflecting variance in backgrounds of employees	Inconsistent set of enacted mechanisms that generates outcomes and achievements	Deemphasize reliance on or differences between individuals by not using titles or positions to address individuals;

in an environment of extreme social differences and inequalities	inconsistent with Aravind's mission	Nurturing an atmosphere of achievement by providing constant performance measures without rewarding over-performers; Nurturing a culture of strict task focus and discipline coupled with habit formation through task specialization and training. Consistency in trainings by involving Aravind family members at all times; Strict selection during hiring processes
Increasing the ratio of paying to non-paying patients to increase profits for Aravind or private profits for doctors	Prevents Aravind from achieving its mission;	Committed relations with community partners who drive a constant flow of poor patients; Doctors cannot make patient appointments and need to treat according to demand; Ubiquitous presence of Aravind family doctors stifle opportunities for non-compliance
Patients lack post-surgery compliance	Creates undesired health impacts and dissatisfaction with Aravind; Negative reputation effects decrease demand for both paying and non-paying patients	Take patients through the whole process in the shortest possible timeframe; Design processes such that requirement for multiple visits is eliminated;
Poorest patients do not trust health services and refuse treatment	Inability to access the poor and fulfill Aravind's mission; Inability to grow to a scale commensurate with need levels in India.	Work with trusted community partners to organize field screening camps; Consistency in patient outcomes through stringent quality commitment builds reputation through word-of-mouth effects.
Variation in treatment protocol and deviation from standard processes	Accumulation of errors that threaten Aravind's quality reputation; Uncoordinated action and experimentation that lowers efficiency levels	Highly formalized processes reinforced by constant training; Large scale operations under time pressure limit ability of individuals to reflect on situations (routine/habituation processes); Strict task specialization
Growing levels of alienation, boredom and loss of satisfaction and motivation due to the strict task specialization; Doctors growing fear of de-skilling and loss of marketability due to task specialization	Lower productivity levels; Loss of key resources;	Nurses are uniformly trained and paid across task domains which reduces their negotiating ability (they can be easily replaced) and lowers sentiments of task-inequality; Investment in training and research facilities that provide many contact points with leading international ophthalmologists.
Paying patients posing as poor patients to get free high-quality surgery	Loss of income and inability to maintain or grow Aravind's capacity to scale	By strictly separating hospital facilities used by paying and non-paying patients Aravind triggers a social self-selection where wealthier patients do not want to sit close to poor patients; Provision of better food, individual rooms, air conditioning and other amenities for

		paying patients.
--	--	------------------

REFERENCES

- Ackroyd, S. and Fleetwood, S. (2000). *Realist perspectives on management and organizations*. New York: Routledge.
- Archer, M. (1995). *Realist Social Theory: The Morphogenetic Approach*, Cambridge: Cambridge University Press.
- Archer, M. S. (2002). 'Realism and the problem of agency'. *Journal of Critical Realism*, **5**: 11–20.
- Bhaskar, R. (1975). *A Realist Theory of Science*. Leeds: Leeds Books.
- Bloom, P.N. and Chatterji, A.K. (2009). 'Scaling social entrepreneurial impact.' *California Management Review*, **51**: 114–133.
- Bradach, J. (2003). 'Going to Scale: The Challenge of Replicating Social Programs.' *Stanford Social Innovation Review*, **1**: 18-25.
- Bunge, M. (2006) *Chasing Reality: Strife over Realism*. Toronto: University of Toronto Press.
- Chen, S. and Ravallion, M. (2007). 'Absolute poverty measures for the developing world, 1981-2004.' *Proceedings of the National Academy of Sciences*, **104** (43), 16757–16762.
- Collier, A. (1994) *Critical Realism*. London: Verso.
- Davies, J. B., Sandström, S., Shorrocks, A. and Wolff, E. N. (2008). 'The world distribution of household wealth.' *UNU-WIDER Discussion Paper 2008/03*.
- Demetriou, C. (2009). 'The realist approach to explanatory mechanisms in social science. More than a heuristic?' *Philosophy of the Social Sciences*, **39**: 440-462.
- De Véricourt, F. and Sousa Lobo, M. (2009). 'Resource and revenue management in nonprofit operations.' *Operations Research*, **57**, 1114–28.
- Durand, R. and Vaara, E. (2010). 'Causation, counterfactuals and competitive advantage'. *Strategic Management Journal*, **30**, 1245-64
- Easterly, W. (2002). *The Elusive Quest for Growth*. Cambridge, MA: MIT Press.
- Easton, G. (2000). 'Case research as a method for industrial networks. A realist apologia', in, Ackroyd, S. and Fleetwood, S. (Eds), *Realist perspectives on management and organizations*, 205-19. New York: Routledge.
- Ebrahim, A. and Rangan, K. (2010) 'The limits of nonprofit impact: A Contingency Framework For Measuring Social Performance.' *Harvard Business School Working Paper*, 10-099.

- Fairclough, N. (2005). 'Discourse analysis in organization studies: the case for critical realism.' *Organization Studies*, **26**, 915–939.
- Gerring, J. (2007) 'Review Article: The Mechanistic Worldview: Thinking Inside the Box.' *B.J.Pol.S.*, **37**, 1-19.
- Harré, R. and Madden, E. H. (1975). *Causal Powers: A theory of natural necessity*. Oxford: Blackwell.
- Hedström, P. and Swedberg, R. (Eds) (1998). *Social mechanisms: An analytical approach to social theory*. University Press: Cambridge.
- Hedström, P. and Ylikosky, P. (2010). 'Causal mechanisms in the social sciences.' *Annual Review of Sociology*, **36**, 49-67.
- Hodgson, G.M. (2007). 'Institutions and individuals: Interaction and evolution.' *Organization Studies*, **28**, 95–116
- Johnson, P. and Duberley, J. (2000). *Understanding Management Research: An Introduction to Epistemology*. London: Sage.
- Lawson, T. (1995). 'A Realist perspective on contemporary "Economic Theory"'. *Journal of Economic Issues*, **29**, 1-32
- Lawson, T. (2003) *Reorienting Economics*. London: Routledge.
- Leca, B. and Naccache, P. (2006). 'A critical realist approach to institutional entrepreneurship.' *Organization*, **13**, 627–51.
- Machamer, P., Darden, L. and Craver, C.F. (2000) 'Thinking about mechanisms.' *Philosophy of Science*, **67**, 1-25.
- Mahoney, J. (2001). 'Beyond correlational analysis: Recent innovation in theory and method.' *Sociological Forum*, **16**, 575-593.
- McKelvey B. (2002) 'Model-centered organization science epistemology', in, Baum, J.A.C. (Ed), *The Blackwell Companion to Organizations: 752-780*. Blackwell: Oxford.
- Merton, R. K. (1968) *Social Theory and Social Structure*. Extended Version. New York: The Free Press.
- Morrison, M. and Morgan, M.S. (1999). 'Models as mediating instruments', in, Morgan, M.S. and Morrison, M. (Eds), *Models as Mediators: Perspectives on Natural and Social Science: 10-37*. Cambridge University Press: Cambridge.
- Morgan, M.S. and Morrison, M. (Eds) (1999) *Models as Mediators: Perspectives on Natural and Social Science*, Cambridge: Cambridge University Press.

- Pawson, R. and Tilley, N. (1997) *Realistic Evaluation*. London: Sage.
- Portes, A. (2000). 'The Hidden Abode: Sociology as Analysis of the Unexpected: 1999 Presidential Address.' *American Sociological Review*, **65**, 1-18.
- Rangan, V. K. and Thulasiraj, R. D. (2007). 'Making sight affordable.' *Innovations*, **2**, 35-49.
- Reed, M. (2005). 'Reflections on the realist turn in organisation and management studies.' *Journal of Management Studies*, **42**, 1621- 44.
- Sayer, A. (1992) *Methods in Social Science. A Realist Approach*. 2nd Ed. Routledge, New York.
- Seelos, Ch. (2010) 'Theorizing and Strategizing with Models: Generative Models of Business Models.' *IESE Working Paper*, DI-0857-E.
- Seelos, C. and Mair, J. (2005). 'Social entrepreneurship: Creating new business models to serve the poor.' *Business Horizons*, **48**, 241-46.
- Seelos, C. and Mair, J. (2007). 'Profitable business models and market creation in the context of deep poverty: A strategic view.' *Academy of Management Perspectives*, **21**, 49-63.
- Tilly, C. (1996) 'Invisible Elbow'. *Sociological Forum*, **11**, 589-601.
- Tsang, E.W.K. and Kwan, K. (1989). 'Replication and theory development in organizational science: a critical realist perspective.' *Academy of Management Review*, **24**, 759-80.
- Tsoukas, H. (1989). 'The validity of idiographic research explanations.' *Academy of Management Review*, **14**, 551-61.
- Tsoukas, H. (2000) 'What is management? An outline of a metatheory', in, Ackroyd, S. and Fleetwood, S. (Eds), *Realist Perspectives on Management and Organizations*: 26-44. New York: Routledge.
- Uvin, P. (1995). Fighting hunger at the grassroots: paths to scaling up. *World Development*, **23**, 927-39.
- Uvin, P., Jain, P.S. And Brown, L.D. (2000) 'Think Large and Act Small: Toward a New Paradigm for NGO Scaling Up.' *World Development*, **28**, 1409-1419.
- Van de Ven A. (2007). *Engaged Scholarship: A Guide for Organizational and Social Research*. New York: Oxford University Press.
- Villani, A. (2010). 'Much Ado about Scale.' *Beyond Profit*, accessed at: <http://beyondprofit.com/much-ado-about-scale>

Whitley, R. (1984). 'The scientific status of management research as a practically-oriented social science.' *Journal of Management Studies*, **21**, 369–90.

World Health Organization. (1999). *WHO Fact Sheet N230*. Geneva: World Health Organization.

World Health Organization. (1997). *WHO Fact Sheet N143*. Geneva: World Health Organization.