

# ORGANIZATIONAL CLOSURE COMPETENCIES AND SCALING: A REALIST APPROACH TO THEORIZING SOCIAL ENTERPRISE

Christian Seelos and Johanna Mair

## ABSTRACT

*Purpose – Social entrepreneurs create novel approaches to social problems such as poverty. But scaling these approaches to the dimension of the problem can be a difficult task. In the social enterprise sector, the subject of scaling has become a key dimension of organizational performance. This chapter advances the scholarly literature on the scaling of social enterprises, a literature which is currently in an embryonic stage and characterized by conceptual ambiguity and fragmented perspectives.*

*Methodology/Approach – We engage realist philosophy of science to develop mechanism-based causal explanations of the scaling performance of social enterprises. We also develop a coding scheme to guide systematic empirical analysis and highlight the explanatory power of counterfactuals. Counterfactuals have been largely neglected in empirical research as they represent mechanisms that are enabled but remain unobservable – in a state of suppression or neutralization of their effects.*

---

**Social Entrepreneurship and Research Methods**  
**Research Methodology in Strategy and Management, Volume 9, 147–187**  
**Copyright © 2014 by Emerald Group Publishing Limited**  
**All rights of reproduction in any form reserved**  
**ISSN: 1479-8387/doi:10.1108/S1479-83872014000009014**

*Findings – We question the ability of organizations to “socially engineer” desired outcomes and introduce a new construct – organizational closure competence. Anchored in realism, this construct provides a basis for productive approaches to social engineering. We elaborate on the importance of organizational closure competencies for scaling, derive a series of propositions, and develop ideas for future research and for practice.*

*Research, Practical and Social Implications – Applying a realist lens allows us to add empirical rigor to research on social enterprises and scaling. Our approach constitutes a move from rich narratives to causal models and informs the way we design and evaluate efforts to address important societal challenges.*

*Originality/Value of Chapter – This chapter demonstrates how to operationalize realist philosophy of science for causal explanations of complex social phenomena and better utilize its theoretical and practical value.*

**Keywords:** Social entrepreneurship; scaling; realism; retrodution; causal model; mechanism-based research; counterfactual

“Much Ado about Scale. The buzz word in social enterprise is ‘scale.’” This is the title of an article by [Adrenne Villani \(2010, p. 190\)](#) in “Beyond Profit,” an online magazine that focuses on organizations that target the challenges of poverty. Scholars also attend to scaling as a key indicator of success in the social sector ([Bloom & Chatterji, 2009](#); [Bradach, 2003](#); [Uvin, Jain, & Brown, 2000](#)). But scholarly literature on scaling has not yet generated a consistent perspective about what the phenomenon of scaling actually is or how to study it. A multitude of perspectives seem to drive various discourses around scaling. The paper by [Uvin et al. \(2000, p. 1418\)](#) illustrates this dilemma. The authors claim a “new paradigm” of scaling in which the concept of scaling not only encompasses perspectives of organizational size, but also the “number of spin-offs it created,” the “number of projects that have been taken over by other actors,” “the degree to which it contributed to the social and intellectual diversity of civil society,” “the number of beneficiaries or even the specific policy changes won,” “local capacity built,” “intersectoral contacts developed,” “norms of trust and cooperation strengthened,” and “democratic space and social diversity reinforced.” We are concerned

that this fragmentation of meanings of “scaling” prevents the building of a robust epistemological basis that would enable progress in terms of theoretical validity and practical usefulness.

In this chapter, we explore the following questions: how to make scaling in the social sector an object of systematic scholarly investigation, and how to theorize scaling more productively. In doing this, we engage with the recent call for more mechanism-based explanations in the social sciences (Davis & Marquis, 2005; Elster, 1989; Hedström & Swedberg, 1998; Mahoney, 2001; Mayntz, 2004; Weber, 2006). To overcome the ambiguity that plagues mechanism-based approaches (Mahoney, 2001), we anchor the concept of mechanisms in realist philosophy of science (Bhaskar, 1975; Bunge, 1996, 2006; Collier, 1994; Demetriou, 2009; Gerring, 2010; Manicas, 2006). Realism provides a causal architecture that links actors, mechanisms, and outcomes as a hallmark of generative causality (Bunge, 2006; Harré & Madden, 1975). Realist explanations require a clear explanandum – a focal phenomenon of interest – as a starting point for inquiry (Mayntz, 2004). Thus, we start by developing a definition of organizational scaling that lends itself to causal explanation. We then develop a coding scheme that elaborates a realist meta-theory of what an organization must be like for scaling to be possible. This coding scheme serves as our analytical instrument for empirical investigation of an extreme case of organizational scaling, an eye hospital in India. We conclude with a series of propositions about scaling and social enterprises and derive implications for further research and practice.

## **WHERE THE DISCOURSE OF SCALING MEETS THE REALITY OF POVERTY**

Inequality and poverty persist in important and intolerable dimensions (Chen & Ravallion, 2007; Easterly, 2002). Tackling poverty does not only require innovative approaches but also finding ways to size initiatives up to the scale of the problem. A sense of urgency, excitement, and hope generated by some successful social sector organizations that were able to expand (Seelos & Mair, 2005, 2009) have stimulated a variety of scaling perspectives. However, many of these perspectives remain entirely discursive. For example, telling richer stories about social sector organizations was proposed to expand the meaning of scaling (Uvin et al., 2000). Expanding narratives increases the perceived scale of existing organizations

and thus resonates deeply with the sense of urgency and need for hope in the social sector. Aligned with this trend, social enterprises are experiencing pressure from foundations and grantors to demonstrate more clearly, more quantitatively, and in a much richer way the real scale of their activities and areas of impact, which critically shapes the discourse on social enterprises (Ebrahim & Rangan, 2010).

One consequence of this is a fascination with “success stories” and framing the inability to scale as a sign of failure. Scholarly attempts to systematically identify the success of social enterprises further drive a discourse centering on success factors (Dees, Anderson, & Wei-Skillern, 2004). For example, Bloom and Chatterji (2009) proposed a model which posits that an organization’s success at scaling its 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 authors refer to this as the SCALERS model. The discourse centering on success factors suggests that getting these factors right is fastest and most efficient way to achieve scale in a predictable manner. Unfortunately, changing the way we speak about or account for scaling will probably have little effect on the “reality of poverty.” According to a realist position, important aspects of reality exist independent of our discourses. Taking a *realistic* look challenges the hope and excitement generated by the prevalent discourse around scaling and social enterprises. For example, many microfinance organizations – often mentioned as a prominent example of social enterprise – do not create the type of impact that was hoped or that is claimed. Even experienced and well-established organizations are struggling with growing and expanding their activities outside their home environment. One example is BRAC in Bangladesh, one of the largest and most experienced social enterprises active in education, health, finance, agricultural, and many other domains (Mair, Martí, & Ventresca, 2012). BRAC struggled to scale activities and initiatives that it had developed in Bangladesh to communities in Afghanistan (Seelos, Mair, Battilana, & Dacin, 2011). Scholars have even voiced concerns that many innovative and entrepreneurial organizations seem unable to scale at all (Bradach, 2003; Uvin, 1995).

We propose that taking realist assumptions seriously is critical for advancing theoretical and practical knowledge about scaling and social enterprises. A realist approach allows us to reflect more deeply on how to generate explanations that prioritize causal validity over attractiveness (e.g., richer stories) or convenience (reducing complex phenomena to a few success factors).

## SCALING AND THE LIMITS TO SOCIAL ENGINEERING

DiMaggio (2001) made a strong case for moving from discourses and rich stories to causal models if we are to assess the effectiveness of organizations and inform practice and public policy decisions. Developing a causal model of scaling requires an explicit definition of scaling. We define scaling as an increase in desired organizational outcomes B that are generated by organizational activities A. This definition has a central explanandum “B” – a set of quantifiable and observable outcomes as the starting point for analysis. The explanans “A” comprises an account of the organizational mechanisms that generate these outcomes.

Our definition overcomes three key weaknesses of the scaling perspectives discussed in the previous section. First, in contrast to the calls for richer narratives (as illustrated by Uvin et al., 2000), our definition avoids treating scaling simply as an outcome phenomenon. Richer narratives do not require any changes in what an organization actually does. Establishing causality is thus questionable. Instead, our definition specifies a causal relationship between organizational activities as causes and the levels of desired outcomes as their effects. Second, by narrowing the scope of our definition we avoid the weakness of impact perspectives on scaling as discussed before. Impact often refers to consequences that are more temporally and causally distant from direct organizational action (Ebrahim & Rangan, 2010). Explaining impact requires an account of complex constellations of mechanisms enacted by multiple parties. This severely challenges the validity of causal explanation. Instead, mechanism-based explanations of complex social phenomena such as social enterprise scaling are facilitated by narrowing the scope of investigations (Demetriou, 2009; Gerring, 2007). Chakravarty (2010) argues that narrowing one’s analytical scope increases the validity of one’s claims. We thus chose a narrow definition of scaling that integrates direct and observable consequences of organizational action as explanandum. Third, our definition of scaling is explicitly anchored in a realist philosophy. We thus avoid the shortcoming of only accounting for “success factors” whose presence correlates with desired outcomes. As argued in the next section, accounting for the absence of a particular class of factors, realist counterfactuals, may have even higher explanatory power. The following example illustrates our definition of scaling. The Aravind eye hospital in India was able to scale by providing sight-restoring cataract surgeries to an increasing number of poor patients

in a robust manner over the last 20 years. Over time, it became one of the largest eye hospitals in the world. A key empirical measure of scale for Aravind is the number of patients treated over time (Fig. 1). In addition, Aravind keeps the ratio of surgeries provided to paying (able to pay above costs) to nonpaying (too poor to pay) patients robustly at about 35:65. Thus, surgeries for the poorest grow in line with total surgeries. Regular outcomes of stable positive margins (revenues minus expenses in Fig. 1) indicate growing profits over time. Aravind generates event regularities over time, that is, it enacts a set of mechanisms that generate a robustly growing set of desired organizational outcomes. Thus, our definition of scaling is consistent with the actually existing scaling of Aravind.

Scholars have voiced concerns over the expectations on “social engineering,” the purposeful enactment of robust and desired outcomes of organized action that is implied by our definition of scaling. Merton (1968, p. 122) reminds us that “social life is not as simple as it first seems” and that purposeful action often creates “unintended consequences.” For our scaling perspective this implies that organizations cannot rely on the fact that doing A will create intended outcomes B. Beyond very restricted dimensions of time and space, social action is likely to have unintended consequences that may often be undesirable. When BRAC started operating in Afghanistan, doing similar things that generated desirable outcomes in Bangladesh created undesirable outcomes in Afghanistan. For example,

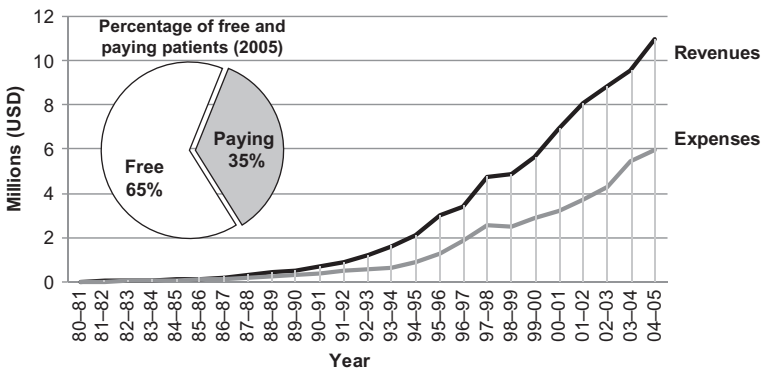


Fig. 1. Revenues and Expenses of Aravind between 1980 and 2005. Source: Data provided by R. D. Thulasiraj, Executive Director of the Lions Aravind Institute of Community Ophthalmology.

microfinance operations failed to generate the desired positive economic outcomes in this particular social context (BRAC management, personal communication, December 2013). Merton (1968) proposes that examining unintended consequences and the circumstances of their workings is more useful for making significant progress in sociological knowledge than identifying “success factors” that are expected to predictably generate intended consequences. Portes (2000) also makes a strong statement for sociology as an 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” and the unpredictability by which it unfolds (Portes, 2000, p. 12). Charles Tilly, considered a pioneer of mechanism-based sociological explanations (Demetriou, 2009) expands this focus on the unexpected by including an awareness of 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 in Tilly’s sense typically do not show up in traditional empirical factor analyses and tend to elude investigations based on a Humean logic of causality as regular associations of cause and effect (i.e., event regularities) (Harré & Madden, 1975; Mahoney, 2001; Sayer, 1992).

Our discussion has three important implications for studying scaling:

1. Scholarly explanations of scaling require an account of the organizational mechanisms that generate outcomes of interest. By insisting on a clearly defined phenomenon that can be objectively established (e.g., numbers of patients treated over time), we can avoid the fallacy of grounding explanations in mere discourses and rich narratives. Important aspects of reality are not dependent on discourses as mechanisms (Bhaskar, 1975; Bunge, 2006; Sayer, 1992).
2. A mechanism-based analysis of scaling challenges us to go beyond an account of observable “success factors.” Realist causal explanations

require asking “what could have been but was not.” This directs attention toward a systematic identification of what we call realist counterfactuals, particularly when they are unobservable. Our definition of scaling implies purposive action and its consequences. Thus, we need to explain how organizational actors find ways to prevent or reduce unintended consequences, particularly those that are not desired. Empirical analysis of observable factors misses this important opportunity for increasing explanatory power.

3. Researchers are confronted with a vast amount of potential explanatory mechanisms in organizations and their environments. We need to develop an analytical tool that guides identification of the relevant mechanisms in a more systematic and transparent manner.

### **HOW REALISM INFORMS EMPIRICAL RESEARCH ON SCALING**

Realist perspectives have been proposed as fruitful avenues for investigation in organization and management studies (Ackroyd & Fleetwood, 2000; Durand & Vaara, 2009; Fairclough, 2005; Johnson & Duberley, 2000; Reed, 2005; Tsang & Kwan, 1989; Tsoukas, 1989; Van de Ven, 2007; Whitley, 1984). Realism provides us with a meta-theory that specifies what must be true for event regularities to occur, that is, a causal model of event regularities. 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.” This is our basis for developing an analytical coding scheme that facilitates empirical research on scaling. In our definition, scaling is an event regularity in the sense that an organization is able robustly to enact a set of mechanisms A to create desired and expected outcomes B over time. Our definition of scaling thus implies a tight coupling between cause and effect: causes can be robustly triggered and robustly generate their expected effects. However, realists claim that this is only possible in strictly controlled experimental systems. Such controlled environments are said to provide causal closure – a context that enables a tight coupling between cause and effect (Sayer, 1992; Tsoukas, 2000). However, in the real world there are multiple possible correlations between a cause and an effect. Realist scholars are pessimistic about generating causal closure conditions in social systems as is implied in our definition of scaling. Bhaskar (1975, p. 25, xxxi) grounded



his realism in a social world characterized by open systems “where causal laws are out of phase with patterns of events and experiences” and where “a constant conjunction is no more a necessary than a sufficient condition for a causal law.” Archer (1998, p. 190) is concerned that “Even in isolated environments, the nature of humans means that “closure” cannot be achieved.” However, Tsoukas (2000, p. 40) suggests there is an opportunity to escape the indeterminacy of open systems: “In other words, management must create conditions of organizational 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.” Thus, realism offers a theoretical architecture that facilitates analytical approaches to empirically examine organizational phenomena that imply a robust coupling between causes and desired and expected effects – conditions of quasi-closure.

## **DEVELOPMENT OF A CODING SCHEME**

The principal mode of inference in realist mechanism-based explanation is retrodution which explains a phenomenon in terms of the mechanisms that possibly, plausibly, or actually generated it (Machamer, Darden, & Craver, 2000; 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. 28). To enable a transparent and systematic process of retrodution, we develop a generic coding scheme for empirical investigations of scaling. The coding scheme is an analytical tool. It enables focused and productive engagement with a focal phenomenon. The outcome of this is a causal model that accounts for the mechanisms that generate observed organizational scaling performance. An important part of establishing the validity of causal models is to specify the relations of the model to (i) a study’s principal analytical interest or research question, (ii) to appropriate theory, and (iii) to ontology, or how a model’s concepts relate to reality (Seelos, 2014). This needs to be reflected in our coding scheme. We have already specified our analytical perspective on scaling as an event regularity. Accordingly, we engaged realist meta-theory because it contributes an important conceptual architecture underlying event regularities to our coding scheme. This includes the concepts of enabled-, triggered-, desired-, and undesired mechanisms as described

below. To specify our model–ontology relation, we engage the realist perspective of Bunge (2004, 2006) who defines the minimal required set of specifications for modeling a material system (such as an organization) as: “constituents,” “structure,” “mechanisms,” and “environments.” Bunge’s perspective is seen as a viable alternative that avoids the shortcomings of both methodological individualism and holism (Bunge, 2004; Pickel, 2004). To develop our coding scheme, we make the following adjustments:

1. We limit “constituents” to human actors and their individual characteristics that are not primarily due to their particular social relations in an organization at any given point in time. For this study, we ignore non-human objects because we do not expect them to create systematic unpredictable variance that would undermine the generation of event regularities.
2. “Structure,” following Hodgson (2007) and Tsoukas (1989), refers to the set of social relations in an organization that have both enabling and constraining effects on the generation of mechanisms. For example, the particular relation of boss–employee enables in the boss the mechanism to fire the employee but not vice versa.
3. Because mechanisms are enabled by the particular individual characteristics of actors and their social relations, the variables “constituents” and “structure” together determine the system potential, that is, the set of enabled mechanisms (Bhaskar, 1975; Bunge, 2006; Harré & Madden, 1975) in the focal phenomenon.
4. “Mechanisms” in our coding scheme refer to the distinct set of enabled 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 outcomes. Because many social outcomes of interest are generated by multiple mechanisms, we bundle related mechanisms into “generative processes” that are comprised of several concurrent mechanisms and/or sequences of mechanisms and outcomes.
5. Because organizations are neither isolated nor independent of their environments, our coding scheme needs to specify the relevant set of internal and contextual actors and the relationships and mechanisms that enable and limit the outcomes an organization can achieve. We thus treat the variable “environment” as the relevant actors, structures, and mechanisms in the task environment of the focal organization.

Following Tsoukas (2000), our definition of scaling implies managerial efforts that generate a robust coupling between actors, mechanisms, and

desired outcomes – the achievement of organizational closure conditions. Realist meta-theory enables us to operationalize organizational closure. A key concept is Harré and Madden's (1975) notion of "enabled mechanisms." "Enabled" means that they are part of the set of properties that define an object, that is, 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. In our illustrative example, Aravind, the mechanism of doing eye surgeries is enabled through proper training of eye doctors or through hiring trained eye surgeons. We code those mechanisms that are causal for generating desired organizational outcomes "desired mechanisms." However, enabled mechanisms constitute a potentiality that may not be realized. In our example, the enabled mechanism of eye surgery also needs to be regularly 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 perform surgeries on a regular basis. Furthermore, realization of the expected outcomes of mechanisms also requires that negatively interfering mechanisms, that is, "undesired mechanisms," are disabled or suppressed. For example, Aravind eye doctors are able and might easily be triggered into leaving for better paid jobs. Managerial effort is required to prevent this undesired outcome that would undermine scaling performance. Alternatively, if the mechanisms of leaving cannot be prevented, its effects need to be neutralized, for example, by building internal capacity to train more eye doctors than are leaving.

The unique identity or constitution of actors as part of the focal organization and its task environment and their relations to each other thus determine the set of enabled desired and undesired mechanisms. Fig. 2 illustrates the way we operationalize realist philosophy of science as a coding scheme. The coding scheme shows that robust observable event regularities, implied in our definition of scaling, are generated by different causes:

- the presence of actors (internal and external to the organization) with the required set of enabled desired mechanisms;
- the positive triggering of desired mechanisms into action on a sustained basis because an absence of triggers would not realize the potential available in an organization;
- the negative triggering to disable, suppress, or neutralize the effects of undesired mechanisms that otherwise could negatively impact outcomes, negatively impact the operation of desired mechanisms, or inhibit the triggering of desired mechanisms into action.

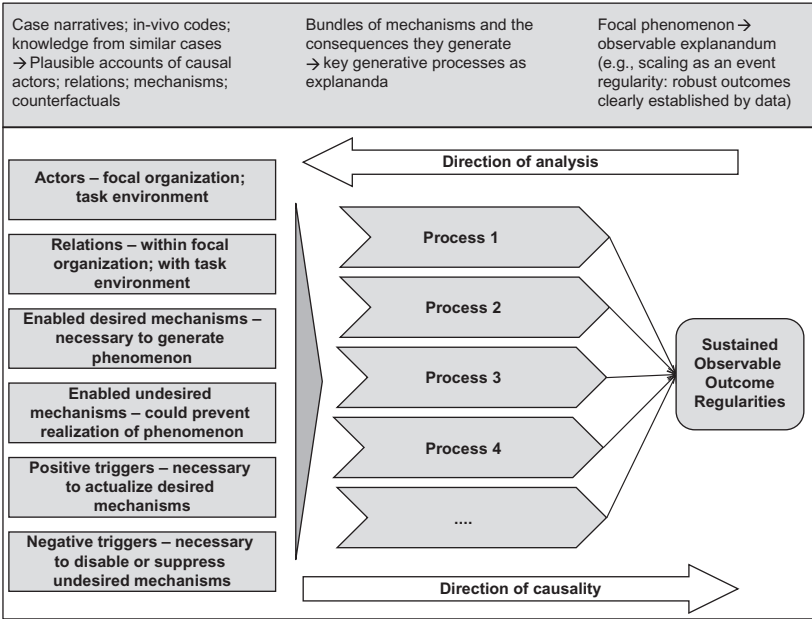


Fig. 2. Coding Scheme.

This coding scheme constitutes an analytical guide for the creation of a causal model that accounts for the mechanisms that generate and thus explain scaling performance as defined in this chapter. The coding scheme helps us to look at the world in a more systematic fashion and guides the analytical process of retrodution. We use the terms “coding scheme” prior to analysis and “causal model” to refer to the outcome of applying the scheme to empirical data and retroduting the generative mechanisms. Thus, while causality works from actors and mechanisms to observable outcomes, the analyst works in the opposite direction. The coding scheme also focuses analytical attention to a particular class of unobservables. In organizations that generate robust levels of specific outcomes, realism implies that undesired mechanisms are actively disabled or suppressed. In other words, undesired mechanisms and their undesired consequences are largely unobservable and their absence is an important “cause” of positive outcomes. We refer to them as realist counterfactuals. By coding for the absence of plausibly enabled and easily triggered undesired mechanisms, such as eye

doctors leaving for better paid jobs, we create analytical attention to realist counterfactuals that we otherwise might not have noticed.

## WORKING WITH THE CODING SCHEME

A key challenge for mechanism-based explanations is the fact that important mechanisms may not be directly observable (Bunge, 2006; Gerring, 2007; Godfrey & Hill, 1995; Hedström & Ylikosky, 2010). To deal with this difficulty, scholars proposed narrowing the scope of investigations to get closer to the micro-foundations of mechanisms (Demetriou, 2009; Gerring, 2007). Practically, this requires searching for mechanisms at their actual levels of instantiation. Central to this approach is that researchers take a critical stance (Leca & Naccache, 2006) and observe actors' actions and practices in addition to the discourses they develop. Investigators can directly capture the reflections of organizational actors in their work environment and their logics of explanation of outcomes as *in vivo* codes. Observing actors in their work environment additionally supports analysis in three important ways: (i) it validates actor logics by critically evaluating fit between discourse and action, (ii) it identifies enabling or constraining actor characteristics and structures that would explain patterned tendencies of observed behavior or habitual tasks that the actors may not reflect upon, and (iii) characteristics of actors and their relations may be too sensitive for interviewees to discuss (e.g., potentially observable power structures or cognitive limitations). Guided by the coding scheme, researchers can 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 effects.

In practice, we construct narratives that comprise the most relevant and plausible direct quotes that illustrate concepts and the linkages between them specified in our coding scheme. Then we conjecture systematically plausible causal links between observable outcome indicators (OI, e.g., profits, high quality, families providing young girls as nurses to Aravind instead of getting them married) and either observable mechanisms (OM) or implied unobservable mechanisms (UM) in the form of analytical OI–OM–UM tables (data not shown). The coding scheme in Fig. 2 also explicitly focuses our attention on realist counterfactuals as an important category of unobservables. One way to code them is to identify events that actors have observed as causes of unintended undesired consequences in

their immediate work environment at some point in time. These are usually recollections in the form of “At the beginning, we tried ... but then we changed to ...”; “Once we hired a doctor who did not ... but we were able to ...”; etc. Realist counterfactuals can also be coded either as plausibly expected events based on relevant theories or observations made by researchers in the same or similar environments and settings or as more general theories that provide potential behavioral patterns that might be operating in a given organization. This requires a deep understanding of the contextual characteristics of actors and relations that constitute a focal phenomenon. Valid explanation of Aravind’s robust scaling performance over time thus requires coding for the mechanisms by which Aravind disables realist counterfactuals (undesired mechanisms specified in Fig. 2), suppresses their realization, or neutralizes their potential undesired consequences.

A next analytical step is to aggregate codes of related mechanisms into smaller sets of key processes. Processes are sets of related micro-mechanisms that create robust higher level organizational outcomes. This analytical procedure increases explanatory power because it creates plausible constellations of several related or mutually reinforcing mechanisms rather than relying on one cause–effect conjecture. For Aravind, we find that a relatively small number of core processes collectively goes a long way in explaining outcome regularities. Demetriou (2009) highlights the need to provide narratives that fill in the blanks, provide context and anecdotal evidence, and strengthen the plausibility of conjectures. We provide in the following section narratives that support the plausibility of conjecturing these processes. The narratives also highlight our central focus on realist counterfactuals. We find that, in the absence of the coding scheme, our analyses tended to be biased toward identifying desired mechanisms, many of which are observable. The coding scheme shifted our attention to counterfactuals and we feel this has greatly increased explanatory power and our understanding of how Aravind ticks. We have summarized the main steps of our analytical approach in Table 1.

## **ANALYSIS OF SCALING AT ARAVIND EYE HOSPITAL<sup>1</sup>**

The basis for this analysis were data gathered until 2007. By 2007, Aravind in India was already the largest group of eye hospitals in the world.

**Table 1.** Main Analytical Steps in Developing Realist Mechanism-Based Causal Explanations.

Analytical Step	Comments
1. Decide on your core explanandum – an aspect of a social phenomenon that can be factually established	Retroduction, the principal realist inference proceeds from effects to an account of their causes. Inability to objectively establish an effect thus systematically challenges the validity of causal explanations.
2. Constrain the scope of your explanandum to facilitate the creation of causal models that validly account for focal phenomena of interest	Narrowing analytical scope enables more valid explanations. Causal models expand and improve over time to further our understanding of broader aspects of focal phenomena. This is an important principle of valid progressive epistemology.
3. Develop a coding scheme and specify the ontological and epistemological aspects that you integrate into your coding scheme to align it with your analytical focus	The coding scheme specifies a causal architecture as a basis for the outcome of analysis – a causal model. Alignment of analytical, epistemological, and ontological dimensions is a prerequisite for causal model validity (for details and references see <a href="#">Seelos, 2014</a> )
4. Gather data by directly interacting with the organization	Requires deep engagement with the phenomenon. The effort is high for both researchers and members of the focal organization. Building of trust is necessary to avoid “story telling” – pleasing researchers with convenient stories to get them “out of the door” and to motivate sharing of counterfactuals.
5. Conjecture plausible causal links between observable outcome indicators (OI) and observable mechanisms (OM) or implied unobservable mechanisms (UM)	Building OI–OM–UM tables is a systematic way of operationalizing retroduction and making it transparent. OMs and UMs can be triangulated by integrating observation, existing theory, accounts of other scholars, organizational members or other stakeholders, and repeated interaction and visits on- and off-site.
6. Identify counterfactuals	Success generated by desired mechanisms implies an absence of counterfactuals – undesired mechanisms. The coding scheme and an explicit analytical focus sharpen attention to this important class of unobservables that have strong explanatory power.

*Table 1. (Continued)*

Analytical Step	Comments
7. Aggregate mechanisms to overall processes	Bundle mechanisms into sets that comprise core processes that generate crucial intermediary outcomes or key characteristics of observed outcomes (volume, quality, financial metrics, etc.). Collectively, these core processes constitute a causal model that ideally goes a long way explaining a phenomenon of interest.
8. Optional – visualize how the individual processes “hang together” into a cause/effect display	This is an additional validity check that establishes the extent to which individual processes are aligned and causally consistent – for example, not canceling each other’s outcomes. The result is a visual causal model (for an example see <a href="#">Seelos, 2014</a> ).
9. Optional – develop propositions	Truthful causal models of focal phenomena increase the validity of inferences from one case to other contexts because the importance of local organizational and contextual factors is well understood.

It pioneered a novel approach to delivering eye surgery for cataracts that integrated free surgeries for the poor as a major strategic objective. In 1976, Dr. Govindappa Venkataswamy (“Dr. V” as he is called at Aravind), a retired ophthalmologist, founded Aravind, an 11-bed eye hospital, in the city of Madurai. Since then Aravind expanded to a group of hospitals that have performed over 300,000 eye surgeries annually and provided eye care services to more than two million outpatients. We highlight a number of mechanisms that generate Aravind’s ability to robustly increase scale. Guided by our coding scheme, we clustered key mechanisms into five core processes. These processes explain how robust desired outcomes are generated at an increasing scale. We thus label them generative processes. They are constituted by bundles of desired and regularly triggered mechanisms and the ways in which potentially undesired mechanisms are disabled or suppressed. They thus constitute a causal model that goes a long way in explaining Aravind’s scaling performance. We also provide a summary of the main counterfactuals, their potential consequences, and the ways in which Aravind disables or suppresses them in [Appendix B](#).



*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 or Sekem in Egypt exemplify this tendency (Seelos & Mair, 2007).

Expansion of scope increases the set of organizational and external actors and their 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 – the coupling of causes and desired effects as discussed in the previous section. 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 decision and public commitment of Aravind’s board to a new “stretch goal”: to build the capacity to provide one million cataract surgeries annually. 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 are cataract surgeries, while the other third comprises 11 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 are essential to the success of Aravind’s mission. As summarized in [Appendix B](#), many counterfactuals could undermine Aravind’s scaling performance. 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). The level of surgeries of the other hospitals is 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 comparison, Aravind’s ability to deliver surgeries at increasing scale is extremely robust (see [Fig. 1](#)). We identified three main 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 the establishment of a rigorous selection process for various actors.

The reputation and trust that Aravind has built over the last 30 years is reflected in 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 14 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, that is, a combination of skill focused and values based training that enables and triggers the set of 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 could severely limit the ability of key personnel to make time 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. However, 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 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 or minimize potential undesired mechanisms, 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 a de-emphasis on status and dependency 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 spans across all task domains as an important way to enable and trigger desired mechanisms and 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 65% 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 nonpaying 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 nonpaying patients (e.g., 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 nonpaying to paying patients at 65:35 is a highly relevant characteristic of Aravind's scaling performance. Cataract services in India's public hospitals are financially out of reach of the poor. However, being poor and blind is a dramatic burden exacerbated by 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 toward 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 postsurgery 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. Noncompliance 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 nonpaying patients to the detriment of the 65:35 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 post-surgical treatment (Rangan & 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, which resembles 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 compromise regularities in outcomes.

The extreme task specialization of surgeons creates concerns over deskilling and loss of marketability among doctors. This could trigger the undesired mechanism of highly trained eye doctors leaving for better paid jobs. India has few eye doctors and they are in high 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. Plus, 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). 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, that is, 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 the number of surgeries performed by each doctor are put on a wall at the end of each day. While the name of the doctor is not openly revealed, it is (semi) private information to inform the doctor about their 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, for example, allocating scarce resources to doctors’ pet projects.

#### *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. Furthermore, profitability is an essential means to realize Aravind’s mission and 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 postsurgical service for paying patients. In addition to nurturing Aravind’s reputation for high-quality surgery, the authenticity generated by sustaining an ability to treat poor patients 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 and most productive eye surgeons of the world – a crucial desired mechanism for lowering costs and improving patient outcomes.

### **THEORETICAL AND EMPIRICAL RELEVANCE OF A REALIST APPROACH TO RESEARCH ON SCALING AND SOCIAL ENTERPRISE**

In this chapter, we developed a coding scheme based on a realist meta-theory and applied it to examine the scaling efforts of the social enterprise Aravind. Following the nomenclature of our scaling definition, we provided objective data to establish increasing levels of outcomes “B” and elaborated the causally effective set of organizational activities “A,” the key mechanisms and processes that generate these outcomes. This constitutes a causal model that explains important aspects of scaling at Aravind. Our objective was to demonstrate the theoretical and empirical usefulness of this approach for research on social enterprise. We argue that the causal architecture provided by our approach potentially generates causally valid and insightful explanations of the nature and performance of scaling efforts. We deliberately used a narrow definition of scaling and our empirical analysis is limited to one – instrumental – case, which imposes limits in terms of generalizability. Yet, the outlined coding scheme and the in-depth analysis of the Aravind Eye Hospital case allowed us (1) to clarify the conceptual contours of *organizational closure competencies*, an important theoretical and empirical construct explaining scaling performance; (2) to add more analytical rigor to understand different ways of scaling; and (3) to derive meaningful implications for practice and future research. We elaborate on these points in detail in the remainder of this chapter.

#### *Organizational Closure Competencies as a Central Construct Explaining Scaling Performance*

A tight coupling of activities A and outcomes B resembles a closure condition (Bhaskar, 1975; Sayer, 1992). In the sciences, “closure” is achieved by

setting up controlled experiments so that a robust coupling between objects, mechanisms, and outcomes can be observed. In such a situation, one can learn which action A will predictably create outcome B. However, social phenomena such as organizations occur in relatively open systems (Archer, 1995; Bhaskar, 1975). Thus, 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 & Tilley, 1997, p. 60). We thus conceptualize organizational closure as a required organizational characteristic for sustained outcome regularities. Our coding scheme in Fig. 2 is a basis for understanding organizational closure – a specification of the generic organizational conditions that must be true for scaling as defined in this study to be possible. Organizational closure implies knowledge about the required set of desired mechanisms and how they are enabled and actualized by particular characteristics and relations of actors within the organization and its task environment. Organizational closure also requires knowledge about potential undesired mechanisms and the ways in which they can be disabled or suppressed. This implies a distinction 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. We therefore put forward the concept of "organizational closure competencies" as an important concept for organizational scaling.

**Definitional Statement.** Organizational closure competencies increase degrees of organizational closure. They constitute an organization's ability to generate a robust coupling between causes and effects. More operationally, closure competencies are comprised of organizational structures and processes that enable and trigger desired mechanisms and disable or suppress undesired mechanisms in a robust manner.

Our analytical approach together with the introduction of organizational closure competencies as an anchoring construct advances existing scholarly discussions on several topics. These include scaling approaches and performance of social enterprises (Dees et al., 2004) and ways of scaling discussed in the broader management literature (Sutton & Rao, 2014), such as replication (Bradach, 1997) and knowledge transfer (Szulanski & Jensen, 2006). In the following section, we develop a series of propositions that advance existing discussions and inform and guide future research.

The propositions enable us to differentiate between four modes of scaling: (i) scaling as increased productivity, (ii) scaling as increased size, (iii) scaling as replication, and (iv) scaling as knowledge transfer.

In a context of deep poverty, most types of organizational resources are scarce. Both desired and undesired mechanisms therefore use scarce resources. In Aravind, this would include scarce eye surgeons or community partners that organize access to the rural poor. Eye surgeons for example may choose to engage in clinical research instead of doing surgeries or spend more time in a private practice instead of working at Aravind. Both constitute undesired mechanisms for scaling at Aravind. If undesired mechanisms can be suppressed and more desired mechanisms can be triggered on a sustained basis, resource productivity will increase. Thus, more organizational output can be generated from the same amounts of resources.

**Proposition 1.** Increasing the ratio of desired to undesired mechanisms generates higher degrees of organizational closure and thus increases resource productivity.

Deeper knowledge about how to enable and trigger desired mechanisms and how to suppress undesired mechanisms is a prerequisite for an organizational competence of achieving higher degrees of closure. However, higher degrees of closure better resemble conditions for controlled experimentation. Controlled experimentation enables better isolation of the effects of newly introduced mechanisms and thus facilitates systematic learning. This strengthens the quality of organizational knowledge about cause and effect relations. Productivity as a key driver of scaling performance thus improves by (i) replacing desired mechanisms with substitutes that are more efficient and effective; (ii) learning about potential undesired mechanisms and how to deal with them; and (iii) identifying and eliminating functionally neutral mechanisms that use scarce resources (organizational slack). Importantly, knowledge as an isolated variable does not necessarily explain performance. We argue that the relation between knowledge and closure is an iterative process with significant explanatory power for scaling performance (also see Proposition 6 below).

**Proposition 2.** Organizational closure competencies increase productivity levels through a positive feedback mechanism between the quality of knowledge and higher degrees of closure.

Productivity increases have particular effects on performance levels in service models where paying customers subsidize the service of nonpaying

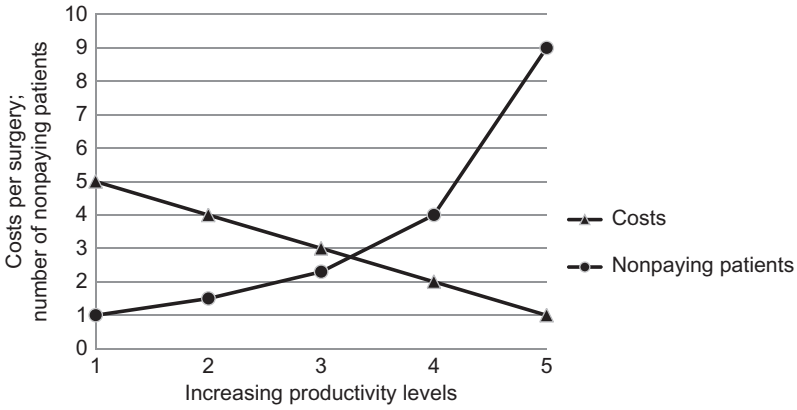


Fig. 3. Relationship between Increasing Productivity Levels of Performing Eye Surgery (Indicated by Lower Costs per Surgery) and Ability to Treat Nonpaying Patients.

customers. Given the constraint of financial self-sufficiency as an organizational objective, as is the case for Aravind, increasing the ability to serve the poor 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 Fig. 3). Using Aravind as an example, we assume that paying patients pay 10 units for a cataract surgery. Initially, productivity is low (productivity level “1” in Fig. 3) at a cost per surgery of five units (with five units of profit available to subsidize one free surgery). Increases in productivity create lower marginal costs per surgery. In Fig. 3, at the highest productivity level “5” marginal costs have decreased to one unit per surgery which enables nine free surgeries for every paid surgery.

**Proposition 3.** For social sector organizations that employ cross-subsidized models, a linear increase in productivity generates nonlinear capacity growth to service nonpaying customers.

*Scaling as Increased Size*

An obvious scaling mechanism is the addition of resources to an organization. Organizational closure improves the ability to identify the key resources

that enable desired mechanisms. Adding and integrating key resources thus generates higher scaling performance. However, adding resources increases organizational size and complexity that challenges maintaining 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*” (Aravind Director). Instead of growing the initial Aravind hospital beyond its current size, Aravind has decided to build four hospitals at almost equal sizes and productivity levels.

**Proposition 4.** Adding key resources that generate desired mechanisms to an organization has a positive effect on scaling until a point where marginal loss of closure due to organizational complexity equals marginal gain from additional resources.

#### *Scaling as Replication*

Replication, that is, creating and operating several identical outlets for producing a product or service, has been proposed as an important way of scaling in the management literature (Winter & Szulanski, 2001). Proposition 4 implies 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 because all resources, not just key resources, are required to replicate an organization. Proposition 2 implies that higher levels of organizational closure enables replication because of the higher ratio of productive to nonproductive resources. Furthermore, proposition 2 implies that higher degrees of closure facilitate the replacement of established desired mechanisms by better substitutes as an outcome of organizational learning. Replication creates organizations with comparable organizational context in which new mechanisms are more likely to work as expected. This facilitates diffusion of new desired mechanisms. Aravind’s branch hospitals regularly exchange comprehensive reports on finance, surgery performance, and quality and engage in sharing and diffusing best practices.

**Proposition 5.** Given resource constraints, higher levels of productivity increase scaling performance through replication.

*Scaling as Knowledge Transfer*

Knowledge transfer to other organizations is a potential mechanism for scaling. Propositions 1 and 2 imply that knowledge transfer is unlikely to increase scaling performance if the receiving organization has lower levels of organizational closure competencies. Knowledge is a prerequisite for closure competencies but may in itself be ineffective unless it was developed in an iterative process as formalized in Proposition 2. Aravind has an internal consulting unit that has advised hundreds of hospitals mostly in developing countries. However, sustained improvements are not always achieved and few hospitals ever achieved anything close to Aravind's productivity levels. One of its innovative practices of managing the setup of new hospitals for three years and then handing them over to a partner organization that has been trained by Aravind has been abandoned because the quality and efficiency levels could not be maintained posthandover.

**Proposition 6.** Knowledge transfer from an organization with high closure competencies to an organization with lower closure competencies will generate lower levels of scaling performance in the receiving organization than in the original organization.

## LOOKING AHEAD

Despite the growing interest among practitioners in the subject of social enterprises, the academic literature on organizational scaling is in an embryonic stage. The objective of this chapter was to infuse research on social enterprises with analytical rigor by identifying causal factors of scaling performance. We did so by reviving mechanism-based approaches and by taking more explicit stands on questions raised by philosophy of science. The approach presented in this chapter allowed us to clarify the ontological and epistemological identity of mechanisms and thus remove both conceptual ambiguity and empirical hurdles for mechanism-based research on social enterprises. Although Tsoukas (1989) has argued for the epistemological validity of idiographic case research from a realist tradition, realist philosophy has received limited explicit attention in research on strategy and organizations in general and the social enterprise specifically. Similarly, while organizational theorists have repeatedly emphasized

the need for problem-driven research based on mechanism-based theorizing (Anderson et al., 2006; Davis & Marquis, 2005; Weber, 2006), social enterprise research has not explored this potential. We argue in this chapter that epistemological and ontological foundations derived from a realist philosophy of science provide a more granular understanding of how “nuts and bolts” based research (Elster, 1989) can be used for both theorizing and empirical work. We consider this approach to be particularly useful for research on social enterprises as it allows us better to understand the role of organizations and organizing in society and at the same time to highlight the practical relevance of theory and research (Suddaby, 2012).

It is our hope that this chapter encourages and inspires future research. Three potential avenues for future research are particularly promising. First, while it centers on scaling performance, this chapter provides a stimulus for more ambitious approaches to study performance of social enterprises. The analytical rigor implied in our approach allows scholars to go beyond statistical inference, rethink causality, and develop a more holistic assessment of the activities of social enterprises and measures of success (DiMaggio, 2001; Ebrahim & Rangan, 2010). Second, while our analytical efforts are limited to one case, future research can develop research designs that allow for comparison of social enterprises operating in similar or different issue domains and geographies. Comparative research is particularly important to more explicitly isolate and clarify the relationship between institutional arrangement and social enterprises (Mair & Martí, 2006, 2009; Seelos et al., 2011) and to inform ongoing efforts to theorize social enterprises (Battilana & Lee, 2014; Dacin, Dacin, & Tracey, 2011; Santos, 2012). Third, our chapter puts a first stake in the ground for deliberate efforts to overcome academic silos and foster more productive exchanges between philosophy of science and disciplinary perspectives. We believe that doing so by studying organizations that go beyond existing ways of doing things – social enterprises – is both powerful and important. Finally, our chapter hopes to rekindle momentum among – especially junior – scholars for doing research that meets high standards of both rigor and relevance.

## NOTE

1. See [Appendix A](#) for a description of interview data collected for this study.



## ACKNOWLEDGMENT

We are grateful to the Rockefeller Foundation and the Center for Philanthropy and Civil Society at Stanford University for their continuous support of our research in the areas of innovation and scaling in social sector organizations.

## REFERENCES

- Ackroyd, S., & Fleetwood, S. (2000). *Realist perspectives on management and organisations*. London: Routledge.
- Anderson, P. J. J., Blatt, R., Christianson, M. K., Grant, A. M., Marquis, C., Neuman, E. J., ... Sutcliffe, K. M. (2006). Understanding mechanisms in organizational research. *Journal of Management Inquiry*, 15(2), 102–113.
- Archer, M. (1995). *Realist social theory: The morphogenetic approach*. Cambridge: Cambridge University Press.
- Archer, M. (1998). Introduction: Realism in the social sciences. In M. Archer, R. Bhaskar, A. Collier, T. Lawson, & A. Norrie (Eds.), *Critical realism: Essential readings* (pp. 189–205). London: Routledge.
- Battilana, J., & Lee, M. (2014). Advancing research on hybrid organizing – Insights from the study of social enterprises. *The Academy of Management Annals*, 8(1), 397–441.
- Bhaskar, R. (1975). *A realist theory of science*. Leeds: Leeds Books.
- Bloom, P., & Chatterji, A. (2009). Scaling social entrepreneurial impact. *California Management Review*, 51(3), 114–133.
- Bradach, J. (2003). Going to scale: The challenge of replicating social programs. *Stanford Social Innovation Review*, 1(1), 18–25.
- Bradach, J. L. (1997). Using the plural form in the management of restaurant chains. *Administrative Science Quarterly*, 42(2), 276–303.
- Bunge, M. (1996). *Finding philosophy in social science*. Yale: University Press.
- Bunge, M. (2004). How does it work? The search for explanatory mechanisms. *Philosophy of the Social Sciences*, 34(2), 182–210.
- Bunge, M. (2006). *Chasing reality: Strife over realism*. Toronto, ON: University of Toronto Press.
- Chakravartty, A. (2010). *A metaphysics for scientific realism*. Cambridge: Cambridge University Press.
- Chen, S., & 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.
- Dacin, M. T., Dacin, P. A., & Tracey, P. (2011). Social entrepreneurship: A critique and future directions. *Organization Science*, 22(5), 1203–1213.
- Davis, G. F., & Marquis, C. (2005). Prospects for organization theory in the early twenty-first century: Institutional fields and mechanisms. *Organization Science*, 16(4), 332–343.
- Dees, J. G., Anderson, B. B., & Wei-Skillern, J. (2004). Scaling social impact. *Stanford Social Innovation Review*, 1(4), 24–32.

- Demetriou, C. (2009). The realist approach to explanatory mechanisms in social science. More than a heuristic? *Philosophy of the Social Sciences*, 39(3), 440–462.
- DiMaggio, P. (2001). Measuring the impact of the nonprofit sector on society is probably impossible but possibly useful: A sociological perspective. In P. Flynn & V. A. Hodgkinson (Eds.), *Measuring the impact of the nonprofit sector* (pp. 247–272). New York, NY: Kluwer Academic/Plenum Publishers.
- Durand, R., & Vaara, E. (2009). Causation, counterfactuals and competitive advantage. *Strategic Management Journal*, 30(12), 1245–1264.
- Easterly, W. (2002). *The elusive quest for growth*. Cambridge, MA: MIT Press.
- Ebrahim, A., & Rangan, K. (2010). *The limits of nonprofit impact: A contingency framework for measuring social performance*. HBS Working Paper Number 10-099. Harvard Business School.
- Elster, J. (1989). *Nuts and bolts for the social sciences*. Cambridge: Cambridge University Press.
- Fairclough, N. (2005). Discourse analysis in organization studies: The case for critical realism. *Organization Studies*, 26(6), 915–939.
- Gerring, J. (2007). Review article: The mechanistic worldview: Thinking inside the box. *British Journal of Political Science*, 37(1), 1–19.
- Gerring, J. (2010). Causal mechanisms: Yes, but .... *Comparative Political Studies*, 43(11), 1499–1526.
- Godfrey, P. C., & Hill, C. W. L. (1995). The problem of unobservables in strategic management research. *Strategic Management Journal*, 16(7), 519–533.
- Harré, R., & Madden, E. H. (1975). *Causal powers: A theory of natural necessity*. Oxford: Blackwell.
- Hedström, P., & Swedberg, R. (Eds.). (1998). *Social mechanisms: An analytical approach to social theory*. Cambridge: Cambridge University Press.
- Hedström, P., & Ylikoski, P. (2010). Causal mechanisms in the social sciences. *Annual Review of Sociology*, 36(1), 49–67.
- Hodgson, G. M. (2007). Institutions and individuals: Interaction and evolution. *Organization Studies*, 28(1), 95–116.
- Johnson, P., & Duberley, J. (2000). *Understanding management research: An introduction to epistemology*. London: Sage.
- Lawson, T. (2003). *Reorienting economics*. London: Routledge.
- Leca, B., & Naccache, P. (2006). A critical realist approach to institutional entrepreneurship. *Organization Studies*, 13(5), 627–651.
- Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 67(1), 1–25.
- Mahoney, J. (2001). Beyond correlational analysis: Recent innovation in theory and method. *Sociological Forum*, 16(3), 575–593.
- Mair, J., & Martí, I. (2006). Social entrepreneurship research: A source of explanation, prediction, and delight. *Journal of World Business*, 41(1), 36–44.
- Mair, J., & Martí, I. (2009). Entrepreneurship in and around institutional voids: A case study from Bangladesh. *Journal of Business Venturing*, 24(5), 419–435.
- Mair, J., Martí, I., & Ventresca, M. J. (2012). Building inclusive markets in rural Bangladesh: How intermediaries work institutional voids. *Academy of Management Journal*, 55(4), 819–850.
- Manicas, P. (2006). *A realist philosophy of social science. Explanation and understanding*. Cambridge: Cambridge University Press.

- Mayntz, R. (2004). Mechanisms in the analysis of social macro-phenomena. *Philosophy of the Social Sciences*, 34(2), 237–259.
- Merton, R. K. (1968). *Social theory and social structure. Extended version*. New York, NY: Free Press.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London: Sage.
- Pickel, A. (2004). Systems and mechanisms: A symposium on Mario Bunge's philosophy of social science. *Philosophy of the Social Sciences*, 34(2), 169–181.
- Portes, A. (2000). The hidden abode: Sociology as analysis of the unexpected: 1999 presidential address. *American Sociological Review*, 65(1), 1–18.
- Rangan, V. K., & Thulasiraj, R. D. (2007). Making sight affordable. *Innovations*, 2(4), 35–49.
- Reed, M. (2005). Reflections on the realist turn in organisation and management studies. *Journal of Management Studies*, 42(8), 1621–1644.
- Santos, F. (2012). A positive theory of social entrepreneurship. *Journal of Business Ethics*, 111(3), 335–351.
- Sayer, A. (1992). *Methods in social science. A realist approach* (2nd ed.). New York, NY: Routledge.
- Seelos, C. (2014). Theorizing and strategizing with models: Generative models of social enterprises. *International Journal of Entrepreneurial Venturing*, 6(1), 6–21.
- Seelos, C., & Mair, J. (2005). Social entrepreneurship: Creating new business models to serve the poor. *Business Horizons*, 48(3), 241–246.
- Seelos, C., & Mair, J. (2007). How social entrepreneurs enable human, social, and economic development. In V. Rangan, J. Quelch, G. Herrero, & B. Barton (Eds.), *Business solutions for the global poor: Creating social and economic value* (pp. 271–278). San Francisco, CA: Jossey-Bass.
- Seelos, C., & Mair, J. (2009). Hope for sustainable development: How social entrepreneurs make it happen. In R. Ziegler (Ed.), *An introduction to social entrepreneurship: Voices, preconditions, contexts* (pp. 228–246). Cheltenham, UK: Elgar.
- Seelos, C., Mair, J., Battilana, J., & Dacin, M. T. (2011). The embeddedness of social entrepreneurship: Understanding variation across local communities. In C. Marquis, M. Lounsbury, & R. Greenwood (Eds.), *Communities and organizations* (Vol. 33, pp. 333–363). Research in the Sociology of Organizations. Bingley, UK: Emerald Group Publishing Limited.
- Suddaby, R. (2012). Editor's comments. *Academy of Management Review*, 37(1), 6–9.
- Sutton, R. I., & Rao, H. (2014). *Scaling up excellence*. New York, NY: Random House.
- Szulanski, G., & Jensen, R. J. (2006). Presumptive adaptation and the effectiveness of knowledge transfer. *Strategic Management Journal*, 27(10), 937–957.
- Tilly, C. (1996). Invisible elbow. *Sociological Forum*, 11(4), 589–601.
- Tsang, E. W. K., & Kwan, K. (1989). Replication and theory development in organizational science: A critical realist perspective. *Academy of Management Review*, 24(4), 759–780.
- Tsoukas, H. (1989). The validity of idiographic research explanations. *Academy of Management Review*, 14(4), 551–561.
- Tsoukas, H. (2000). What is management? An outline of a metatheory. In S. Ackroyd & S. Fleetwood (Eds.), *Realist perspectives on management and organizations* (pp. 26–44). New York, NY: Routledge.
- Uvin, P. (1995). Fighting hunger at the grassroots: Paths to scaling up. *World Development*, 23(6), 927–939.

- Uvin, P., Jain, P. S., & Brown, L. D. (2000). Think large and act small: Toward a new paradigm for NGO scaling up. *World Development*, 28(8), 1409–1419.
- Van de Ven, A. (2007). *Engaged scholarship: A guide for organizational and social research*. New York, NY: Oxford University Press.
- Villani, A. (2010). Much Ado about Scale. *Beyond Profit*. Retrieved from <http://beyondprofit.com/much-ado-about-scale>
- Weber, K. (2006). From nuts and bolts to toolkits: Theorizing with mechanisms. *Journal of Management Inquiry*, 15(2), 119–123.
- Whitley, R. (1984). The scientific status of management research as a practically-oriented social science. *Journal of Management Studies*, 21(4), 369–390.
- Winter, S. G., & Szulanski, G. (2001). Replication as strategy. *Organization Science*, 12(6), 730–743.

## **APPENDIX A: DESCRIPTION OF THE INTERVIEW DATA COLLECTED**

---

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

---

## APPENDIX B: IDENTIFIED COUNTERFACTUALS, THEIR POTENTIAL UNDESIRE 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.	Difficulty prioritizing; increasing organizational complexity and loss of closure; inadequate development of key resources; lower productivity.	Doctors in key positions are relatives of the founder – keep the legacy alive as a group; social conformance pressure suppresses deviance; constant monitoring of narrow output metrics and stretch goals drives focused capacity building.
Lower quality surgeries for nonpaying 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 among 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 nonpaying 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;	Inability to grow output due to low levels of doctors or nurses; Inability to keep a constant pace which generates inefficiencies and service disruption;	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

Aravind trained skilled doctors leave for higher-paid jobs.	Inadequate skills generate low productivity and higher error rates.	constant flow of rural patients to Aravind; Establish their own factory for high-quality intraocular lenses.
Dependency on external funds to subsidize nonpaying patients, for example, 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 in an environment of extreme social differences and inequalities.	Inconsistent set of enacted mechanisms that generates outcomes and achievements inconsistent with Aravind's mission.	Deemphasize reliance on or differences between individuals by not using titles or positions to address individuals; 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 nonpaying 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 noncompliance.

*Appendix B. (Continued)*

Counterfactuals – Potential Undesired Mechanisms	Potential Undesired Consequences	Disabling/Suppressing Mechanisms by Aravind
Patients lack postsurgery compliance.	Creates undesired health impacts and dissatisfaction with Aravind; Negative reputation effects decrease demand for both paying and nonpaying 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 deskilling 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 reduce 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 nonpaying 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.

---