

■ Research Paper

Group Model Building in a Pressure Cooker: A Field Experiment with Public Policy

Lasse Gerrits^{1*} and David Vaandrager²

¹Department of Political Science, Otto-Friedrich Universität Bamberg, Bamberg, Germany

²Department of Public Administration, Erasmus University Rotterdam, Rotterdam, The Netherlands

Group Model Building has established itself as a credible method to uncover the systemic reality of persistent problems in organizations. The contextual nature of the method and the flexibility required during the process mean that it is hard to build up evidence of what works and what does not. One major question is the extent to which the method can be compressed, for example, due to time or budget constraints. We present an experiment, carried out with a public organization, in which we tested two shortcuts. Firstly, we substituted the expert-driven preliminary model for a client-made model. Secondly, we allowed for a homogenous group because enlarging the group would be too time consuming. We defined several factors at the individual level and at the group level as dependent variables. We conclude that the shortcuts had a negative effect on most variables. We therefore recommend against these changes to the basic model. Copyright © 2017 John Wiley & Sons, Ltd.

Keywords Group Model Building; budget cuts; experiment; public organization; public policy

THEORY DEVELOPMENT IN GROUP MODEL BUILDING

Group Model Building or GMB has been around for more than two decades. GMB combines soft systems thinking with a concrete methodology with which groups can develop a shared understanding of the systemic nature of the issues they grapple with. GMB aids in

system improvement and optimization, a somewhat ambiguous goal that breaks down to the following subgoals to achieve (i) alignment of mental models as held by the actors involved with a certain issue or problem; (ii) consensus about a policy or decision to solve things in a certain way; and (iii) support and commitment for the decision (Andersen *et al.*, 1997; see also e.g. Senge, 1990; Vennix *et al.*, 1992; Vennix, 1996).

Whilst GMB has been applied in many instances, it has several persistent problems. As, for example, Richardson and Vennix and others

*Correspondence to: Lasse Gerrits, Department of Political Science, Otto-Friedrich Universität Bamberg, Feldkirchstrasse 21, room 03.71, 96052 Bamberg, Germany.
E-mail: lasse.gerrits@uni-bamberg.de

have pointed out in this journal, GMB is very much an art and less of a science in the sense that it features persistent causal mechanisms. Consequently, it is difficult to build a body of scientific knowledge about what works and what does not. Of course, there are certain things that seem to work across cases: For example, causal loop diagrams are usually helpful in highlighting system traps, a diverse group of participants will lead to more inclusive system representations, and success depends heavily on open communication and transparency among the participants (cf. Andersen *et al.*, 1997; Bérard, 2010; Hovmand *et al.*, 2012; and others).

Such findings give us chunks of scripts that can be assembled to suit the modelling process in particular situations (Hovmand *et al.*, 2012). Given that cases are contextual by definition, a uniform approach or framework that is robust enough to apply to all cases is probably unattainable. The scripted parts offer a reasonable trade-off between proven 'recipes', that is, causal mechanisms, and flexibility. The question then is to what extent these scripted parts can be mixed, omitted, compressed or extended should the situation require. This question is particularly important in those circumstances where there is little room to execute all scripts correctly. How well does GMB work in less than ideal situations whereby there is little room for consensus and time pressure and not everyone is convinced of the approach in the first place? This is the subject of the current paper. Our research question is as follows: Can GMB be compressed with regard to *roles* and *time frames*? To answer this question, we carried out a GMB project with the municipal authorities of Rotterdam (the Netherlands) who were interested in identifying their blind spots with regard to cost-cutting measures and their effects on policy and the city's well-being. We will first give a brief introduction about systems thinking and GBM. This will serve as the foundation for the methodological considerations for the field experiment. We will then explain the setup of the field experiment, discuss the group modelling process and present our findings.

A SYSTEM'S PERSPECTIVE AND GROUP MODEL BUILDING

The popularity of systems thinking in public policies has waxed and waned over the years (Bérard, 2010). It has been criticized for having a supposedly mechanistic worldview, rooted in homeostasis, and for suggesting technocratic solutions (cf. Gerrits, 2012; Turner, 2005) by promising all-inclusive models of society to assist policy-makers with finding the right control parameters to steer society (cf. Klijn and Snellen, 2009). It also appeared to underestimate the importance of reflexive and learning humans in such systems (Flood, 1999a, 1999b). These arguments have been countered by incorporating semiotics and temporal equilibrium states, whilst retaining the valuable insight that societal issues have systemic properties. The central idea is that people working in organizations develop mental models of the systemic properties of a policy issue and will act accordingly, effectively making these models a reality through their actions. Naturally, their preliminary theories of what works and what does not will change due to new practical experiences, effectively updating their knowledge base.

It was Peter Checkland (1981, 2011; Checkland and Howell, 1998) who was particularly instrumental in developing soft-systems methodology as a systems approach rooted in action research. Soft-systems methodology provides both a conceptualization of social systems and a method to map such systems. In his view, the systemic nature of an issue can only be understood through repeated action, that is, experimenting with the assumed causal relationships of the issue, and subsequent reflection on that action because (i) neither researcher nor practitioner can step outside the systemic whole and (ii) the system's workings are best understood through experience with what works and what does not instead of trying to map the system on the basis of theoretical expectations (e.g. Gerrits, 2012; Tsoukas and Hatch, 2001; Wagenaar, 2007). Crucially, Checkland argues that researchers should learn the system's properties by working in it rather than observing from the outside. In adopting Checkland's position, '[...] it becomes

inevitable that a system *definition* turns into a system *methodology*. After all, one can only learn the system through action-and-reflection, a fixed pair of methodological activities'. (Gerrits, 2012: 47, authors' italics). Consequently, system research becomes action research (see, for example, Wagenaar, 2007 for this argument within the context of public policy).

There is a considerable suite of methods for eliciting mental models in a participatory or interactive way in order to highlight the blind spots in an organization or network (Andersen *et al.*, 2007). A particularly useful tool for this aim is GMB. It offers a clear, step-by-step methodology for researching systemic wholes through semiotics and action research, which seems to be associated with positive organizational change (Rouwette and Vennix, 2006). It is useful to summarize the basic stages of GMB (derived from Vennix, 1996) before comparing them to the experiment reported in this article.

Stage 1 Preliminary Model

The method centres on group sessions that tap into the experiences of the participants. However, the process should be started with a preliminary model that is based on *ex ante* interviews and/or documents as an anchor to structure the group sessions. Starting with interviews provides the opportunity to meet group members upfront and establish a rapport. Insufficient knowledge about the participants can be problematic because the model builder will then be more likely to be unaware of potentially problematic dynamics within the group.

Stage 2 Problem Definition and Boundaries

The first group sessions are meant to determine the central problem or purpose. This is a very important phase because the central problem definition guides the effort and safeguards against creating meaningless 'models of everything'. It is important to be very thorough when defining a starting point (i.e. a system's problematic behaviour) from which to embark (cf. Richardson and Midgley, 2007). The issue can be discussed within the group and collectively fine-tuned until it has been defined to the satisfaction of all the group members. Naturally, this

includes a discussion about the issue's boundaries, that is, what should still be considered part of the issue and what is not.

Stage 3 Causes and Consequences

Causes and effects are embedded in a myriad of other causes and effects. This forms the structure from which a system's behaviour emerges. The next step is therefore to look at those causes and effects as identified by the participants. Typically, only the most visible causes will be named. Only by discussing them in-depth will more elements be mentioned. Some will be categorized as direct causes, others as causes of causes with different degrees of directness (first tier, second tier, etc.). As all the relations between the causes and effects are recorded, the overarching model will start to take shape. This is an iterative undertaking that will continue until consensus is achieved. It may take a considerable number of sessions to complete because it requires participants to rethink and talk about their initial mental models. The decision to add or discard causes and effects is guided by the question of whether or not they are seen as having a significant impact on the system's behaviour. This step requires a wide diversity among the participants. If the participants are of a similar mindset, then they are less likely to articulate a more complete model (see e.g. Bérard, 2010 for an extended discussion).

The steps presented here provide a general logic. There are more elaborate scripts and requirements for each step, and there are ample examples of attempts to change parts of the method in order to empower changes to mental models and system performance (e.g. Rouwette and Vennix, 2006; Van Nistelrooij *et al.*, 2015). Reports about such aspects include '[...] the importance of teamwork (Richardson and Andersen, 1995), the identification of pre-defined sets of behaviour in facilitating GMB sessions or "scripts" (Andersen and Richardson, 1997), the sequencing of scripts in the design of GMB interventions (Ackermann *et al.*, 2010), evaluations of GMB effectiveness (Rouwette *et al.*, 2006) and the use of "process maps" as visual tools for designing collaborations (Straus, 2002)'. (Hovmand *et al.*, 2012: 180). However, the picture is far from complete

(cf. Bérard, 2010) and there is still room for experimentation, if only because there is a tendency to only report successful cases (Andersen *et al.*, 1997) and because similar scripts in similar situations still seem to generate different outcomes (Hovmand *et al.*, 2012).

DEVIATIONS FROM THE BASIC PROCEDURE

The field experiment presented here was carried out with the municipal authorities of Rotterdam (the Netherlands). In the wake of the financial crisis, public administration was going through a series of major budget cuts. The second largest, post-industrial city in the Netherlands, Rotterdam has made tremendous efforts to improve the quality of life for its inhabitants. These efforts are expensive and left Rotterdam with a large budget deficit (a record 150 million Euros in 2012). Consequently, it was decided to cut back costs by implementing measures such as severances and the complete abolishment of certain public policies.

Staff at the municipal strategy unit (the client) were anxious about the possibility that the major budget cuts would undermine the city's well-being. Having read Peter Senge's (1990) work, they asked to be trained in systems thinking in order to assess the extent to which budget cuts could be made whilst keeping the urban system stable. The general thrust behind this wish was their belief that civil servants should not just order a system's diagram from a contractor but should learn to think in terms of system dynamics when deciding about budget cuts. In addition to this, the client had to deal with the political calendar that dictated elections in 6 months. It was expected that a change of power would put more pressure on the organization to cut budgets. It was therefore considered crucial to finish the entire process well ahead of the elections so that civil servants would be better able to cope with such pressures. This left us with two points of departure: (i) to let the participants do as much by themselves as possible and (ii) to finish the process within a very limited time span, initially set at 2 months. An argument can be made that these are two contradictory goals, but there are

positive examples of GMB being carried out in short projects (Rouwette, 2012). Based on the literature mentioned in the preceding texts, the client's wishes and our previous experience with GMB (e.g. Vaandrager, Bressers and Gerrits, 2014), we introduced the following changes to the basic process presented in the previous section.

Client-Led Initial Modelling

For the reasons mentioned in the previous section, it is considered good practice to start with a preliminary model made by the researchers. The downside is that it still takes some of the modelling out of the hands of participants. That limits them in learning how to develop a starting point and thwarts their sense of ownership (Scott *et al.*, 2015). In addition, outsiders cannot be expected to develop a correct, albeit limited, preliminary model at short notice (Andersen *et al.*, 2007). However, it appears that the differences between initial models developed by outsider experts or insider participants disappear over the course of the process and have no discernible effect on the outcome (Rouwette, 2012). We therefore decided to let the participants develop a preliminary model under our guidance. We expected that a guided group effort would also lead to a solid starting point.

HOMOGENEITY AMONG PARTICIPANTS

It is generally understood that models improve if the group of participants is heterogeneous because this will bring more diverse perspectives to the table (Berard, 2010). However, it also recognized that a large group with considerable diversity leads to communication problems (Vennix *et al.*, 1992) and the exposure of personal conflicts (Richardson and Andersen, 1995), not to mention that it takes considerable time to accommodate such diversity. A possible way out could be to create subgroups (e.g. Cavana *et al.*, 2007), but that is only feasible when there are ample organizational resources and plenty of time. Because both were scarce, we decided to settle for

a relatively small and homogeneous group of civil servants at the strategic level of the organization. We planned to negate the homogeneity of the group by instructing the participants to collect data from a wide range of sources and to safeguard the diversity of the data (cf. Berard, 2010). We expected that this instruction would introduce more heterogeneity than originally present in the group.

SETUP OF THE FIELD EXPERIMENT

Anderson *et al.* (1997), among others, lamented the lack of theory development and rigorous testing of GMB. Such comments boil down to a desire for approaches to be theory driven, replicated and cumulative. We generally agree with this, except for the aspect of replication. As many authors have pointed out, GMB is heavily context dependent and replication only works in closed, decontextualized settings—something which is simply impossible in complex social situations (cf. Byrne, 2002; Byrne and Callaghan, 2014). Naturally, any attempt at GMB will feature both general characteristics and local idiosyncrasies, something which Anderson *et al.* do recognize. As such, we would not be able to approach anything resembling an ideal experimental setup, that is, the double-blind trial. However, it is inevitable that the research was carried out *in situ*, that is, a field experiment. In this experiment, we tested the two expectations mentioned in the section on Deviation From the Basic Procedure: (i) that the group would be able to identify a solid starting point by themselves and (ii) that the homogeneity of the group could be off-set by asking the participants to collect diverse data from a wide range of sources.

The field experiment was carried out during the autumn of 2013. The 17 primary participants were all staff members of the strategy department of municipal authorities, each representing a specific organizational unit within the broader organization. In the spirit of action research, the authors of this article took on a double role, both guiding the process and reporting on it. To this end, we recorded all sessions by audio means and kept and categorized all written

communications, ranging from short notes to the system diagrams made by the participants. The diagrams were drafted in VENSIM. In addition, we kept diaries in which we not only recorded the details of the process but also our own interpretations and considerations of how the process was developing. These data are summarized in the process description in the succeeding texts.

By what measure would the process be deemed successful or not, thereby confirming or rejecting the two expectations? There is considerable discussion in the literature about the outcomes of GMB. Whilst ‘improving the system’ sounds reasonable from a theoretical perspective, it raises all kinds of methodological questions: After what time span can improvement be expected? How can it be ascertained that the improvement relates directly to the GMB? And how can improvement be assessed when it is often a subjective matter? Rouwette and Vennix (2006) suggest that the effects of GMB can be found on several dimensions: individual, group, organization and method. Given the limited time available, we expected it to be highly unlikely that we would witness changes at the level of the whole organization. However, we expected to be able to detect changes at both the individual level (positive response to the process, mental model refinement, commitment and behavioural change) and the group level (increased quality of communication, creation of a shared language, consensus about problem and solution definitions). These items constitute the dependent variables of the experiment.

We will now report the results by following the timeline of the experiment. Given the contextual nature of GMB, a rich description will be provided because this will give the reader insight into the specific conditions of the field experiment and the mechanics at work.

PREPARATORY STAGE

We were first approached by two members of the strategy department for the reasons outlined in the preceding texts. Several options were explored during the initial dialogue. Consequently,

we drafted a project proposal, including a preliminary central question and a number of possible research methods. The central question was as follows: How many budget cuts could the municipal authorities implement without ceasing to function properly in policy areas that were deemed critical? Given the relatively broad scope and vague wording of the question, it was necessary to focus on a specific policy field. Waste and cleaning policies were suggested as a suitable testing ground. In an age of austerity, such policies may be perceived as being too costly and city councils may find it tempting to allocate scarce resources to fancier policies. Therefore, the question was narrowed down to this: How much is it possible to cut back on waste and cleaning policies before a downward spiral is triggered?

Two options for answering this question were offered. The first concerned working through different scenarios collaboratively, implying that the organization would both deliver data and be an active participant in the process. The second possibility was that the municipal authority would deliver key figures, which the researchers would process in VENSIM to create a system diagram. The first option was favoured because the second option meant that the civil servants would not learn anything from it.

Next, the researchers were asked to present the plan to the entire strategy unit. The presentation, the content of which had been mutually agreed upon beforehand, contained the central research question, a short introduction to system dynamics, a few examples of similar work conducted by the researchers for other organizations and a proposal for a two-stage approach: a pilot about waste and cleaning policies first and a follow-up covering other policy areas if the pilot were successful.

During the presentation, it became clear that the department was deeply divided about the use and attainability of this project. Proponents repeated the case for mapping the urban system; opponents believed that it would be impossible and something that they would not be able to carry out by themselves. Ultimately, the project was approved, simply because the municipal chief executive was among the supporters.

REDEFINING THE PROJECT

The two strategists wrote a second document in order to outline the exact goals they wanted to reach. The document was also intended to serve as a template to determine the programme and costs. The project's goals were now formulated as follows: (i) to evaluate GMB as a method for assessing the budget cuts by doing a pilot test in the field of waste and cleaning policies; (ii) to develop a better understanding of the relationships, weaknesses and tipping points in that particular policy area; and (iii) to combine both method and results in a written advice to the board of directors. If successful, the project could also be expanded to create a toolkit for municipal strategists to work with. Interestingly, the document explicitly mentioned the desire of the authorities to withdraw from certain societal tasks and to delegate more responsibility to citizens, which was loosely summarized under the header of 'self-organization'. In other words, the municipal strategists had already framed and proposed a solution for a problem that was poorly understood. This desire was motivated by the feeling that the economic downturn would last and that the authorities would never be able to return to the same budgetary level as before. Thus, a seemingly technocratic discussion about budget cuts also became a normative and ideological one.

Another informal meeting between researchers and strategists took place. The strategists demanded that the municipal strategy department should be strongly involved in the whole process, preferably taking over as many research operations as possible in order to master the method themselves. Furthermore, it turned out that the time frame available was much shorter than originally envisaged and there would be time for only two workshops instead of the four that had been initially planned. Following this, we proposed an alternative approach to GMB, as outlined in the section on Deviations From the Basic Procedure.

FIRST SESSION

We developed a short instruction sheet that gave some condensed background information about

system dynamics modelling and offered a toolkit to enable the officials to collect data by themselves. We had to decide on whether or not to use operationalized theoretical concepts. Eventually, we decided to keep those concepts to an absolute minimum and tried to involve the participants by working and gaining experience with the method as early as possible during the sessions.

The first session was used to introduce GMB and to demonstrate how mental models can be reconstructed from interviews. Accepting that the officials would learn quicker through actual practice, they practiced data collection during breakout sessions. The participants were asked to assume the role of either interviewer or respondent. They were told to be as naïve as possible about the topics: Everything the respondent shared had to be regarded as truth. Once the central problem was established during the interview, the interviewers had to ask about the assumed causes of the problem and the causes of those causes. Likewise, they had to ask about the consequences and the consequences of those consequences.

The procedure for this practice had been developed in an earlier research project (authors, 2014) and relied on the experience that the participants would develop a sense of the systemic properties as an issue before being confronted with actual theories about how systems operate. It was expected that the naming of consequences would lead both interviewer and respondent to intuitively see that those effects would feed back into causes, therefore giving them a sense of feedback loops. This actually happened during the workshop: Some of the preliminary models reported after the interviewing exercise already had recognizable feedback structures drawn into them, even though the concept of feedback had not been mentioned by the researchers up until that point. Our general impression at the end of this session was that the audience had accepted the presented ideas and that some basic skills had been instilled, leading to a positive assessment. However, we also recognized that most of the strongest opponents had opted to stay away, if only because they did not want to disturb the process.

DATA COLLECTION

The basic models now instructed the researchers to collect data from interviews to inform the preliminary model. This model would then have to be further developed during a second session. As requested by the client, data collection was relegated to the inexperienced civil servants who had attended the first session. Their data would have to be sent to us, so that we could develop a preliminary model, to be discussed during the second session. It was assumed that as the officials entered the field, they would encounter many different perspectives on the issue of waste and cleaning policies.

The results started coming in 4 days before the second session. Unfortunately, most of the models were of a truly disappointing quality. Few officials had even taken the effort to actually construct a model. Those who had made an attempt had usually done it incorrectly and, it appeared, in a great hurry. The models were also based on very little but homogeneous information. An important reason for this homogeneity was that, contrary to the instructions, most of the civil servants had only interviewed direct colleagues at approximately the same hierarchical level. Even if they did interview outside of their own organization, as two did, they were unable to translate those results into anything other than what they themselves understood to be true. In other words, they only recorded and emphasized those views that confirmed their own. In addition to this, normative and factual statements were mixed.

The resulting composite model did not meet quality standards. Consequently, changes to the process were required in order to save the effort. There was one session left to compensate for the lack of quality. By now, it was clear that the civil servants were unable to generate sufficient quality data. We decided to use the second session to present the preliminary model and to ask for additional information, in other words, to use that session to collect more data. To this end, we focused on those areas in the model where most detail seemed to be lacking as a result of questionable interviewing and rhetorical statements.

THE PRELIMINARY MODEL AND ITS BLIND SPOTS

It is useful to discuss the preliminary model and its shortcomings before reporting on the second session. The data were processed in VENSIM and resulted in the causal diagram shown in Figure 1. Note how the variables and relationships were identified but that the feedback loops were not assigned a quantitative or qualitative value. We identified three main areas where details and precision were lacking, indicated by the three dotted and numbered circles in Figure 1.

The first circle concerned citizens' mentality. When talking about responsible citizens, most respondents were of the opinion that citizens did not have the right mentality for keeping the city clean. A policy measure to change citizens' mentality was therefore deemed necessary. However, it was unclear how mentality could be changed and how that would lead to

increased responsibility regarding waste and cleaning.

The second area featured two questions concerning the role of the municipal authorities. (i) It centred on the current measures and the desired balance in responsibility between citizens and government. If the government was to retreat, then alternative measures were deemed necessary to steer mentality. However, because most existing measures had proven ineffective, there was room for improvement. Furthermore, (ii) there was also a contradiction between expecting citizens' initiatives (under the euphemistic header of 'self-organization') whilst at the same time trying to achieve a very specific sort of behaviour through corrective measures.

The third area focused on the accumulation of waste in public space. It was not specified how it came into being and where the proverbial tap from which waste flows could be tightened.

One final question to the participants was added regarding the model in its entirety: Which

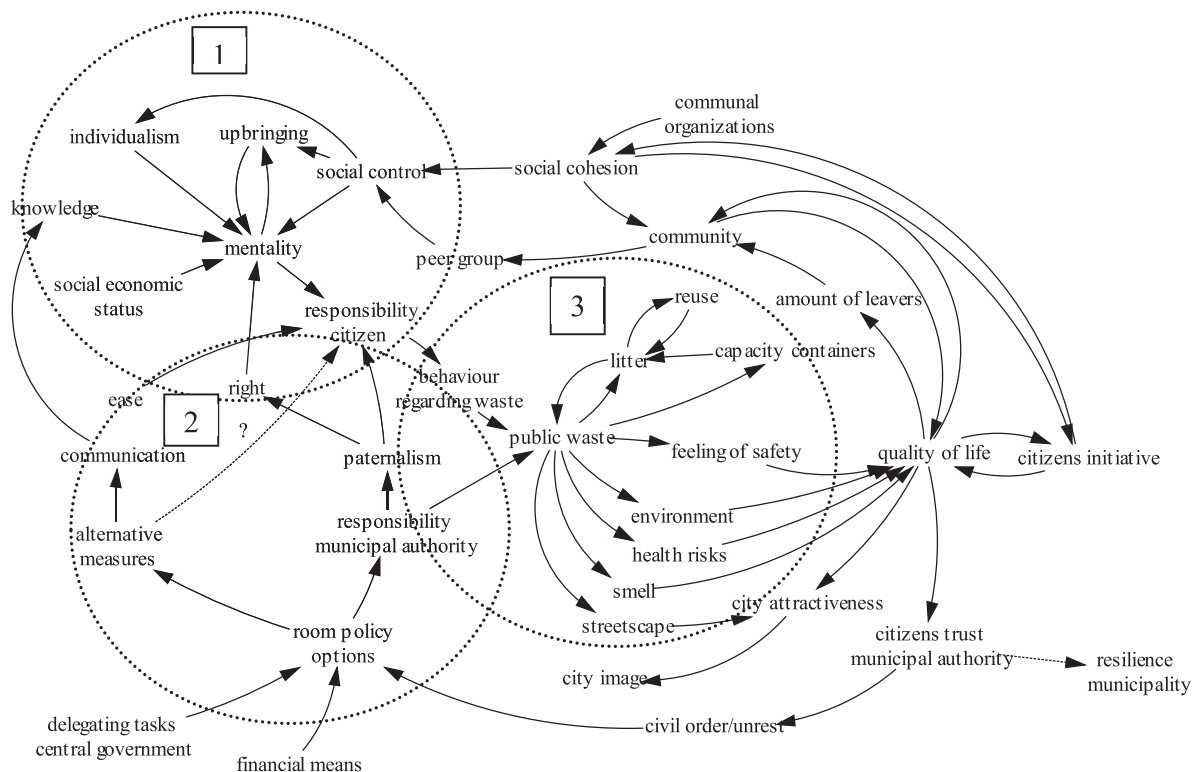


Figure 1 The preliminary model and its blind spots

relations represented by the arrows in the model were based on real observations and which were based on assumptions?

SECOND SESSION

During the second session, the group was divided into five groups representing the five discussion points (as per nominal group technique, see Delbecq *et al.*, 1975). They received a copy of the model and were asked to comment on it, based on the topic assigned to them. They were then asked to report their findings during a plenary discussion. We asked for further details in an attempt to push towards the boundaries of their existing mindsets. At times those barriers were hard to break, with responses ranging from 'we are not in a position to influence those things', to 'political mandates determine everything', to 'we are not in the possession of the right tools' and finally, to 'civilians are simply unwilling'. At the same time, they indicated that they were looking for a holy grail and had accepted that they would never find it because they were not in control. Yet, they did not want to be seen doing nothing about the matter. Our goal soon became to find out why they were not achieving the goals they wanted to achieve and why they felt powerless.

After an extended and sometimes confusing discussion, the session ended. Some participants felt that it was necessary to do one more session in order to obtain the model right. Time and budget restrictions did not really permit this, but the researchers agreed to do one more presentation of a revised version of the model, extended with the findings of this second workshop. The discussions were recorded on audio, which made the effort to integrate the data into a new version of the model somewhat easier. The next section discusses the content of those findings.

RESULTING MODEL

Utilitarian municipal rationale had streamlined conceptualizations of problems and solutions. There was a dominant impulse among the

participants to match the problem definition to the functionally divided departments within the municipal organization. Whilst understandable given their professional background, this pre-selection limited not only the answers that could be sought but also the definitions a problem could be given because the respondents believed that it had to fit within the scope of their departmental tasks. Furthermore, this was done with the clear understanding that the responsible department would never have responsibilities broader than what they were normally accountable for, even if the problem dictated otherwise. At the same time, the intended goal was far broader in scope because the officials did not want a simplistic technical solution for waste and cleaning policies on a lower budget. What the officials really longed for were engaged and responsible citizens.

The questioning during the second session made it clear that the issue of 'responsible citizens' was deemed of far greater importance than a clean city. In the view of the civil servants, responsible citizens were not only supposed to provide a clean public space but also to pick up on a myriad of other tasks, something which was deemed a crucial condition for the coming budget cuts and a smaller municipal organization. Naturally, engaged citizenry was viewed through the narrow sight of the municipal mindset, meaning primarily looking at it through the tools they had to achieve it. A major paradox became visible: 'we are going to challenge complex dynamic interrelated problems with broad impacts by making static, slim and virtually autonomous departments responsible for citizens' self-organization'. As a result, 20 to 30 years of attempting to engage citizens had resulted in a shared uneasy feeling that they could not leave anything up to those citizens. Indeed, they felt that the harder they tried, the less confident they had become.

The institutionalization of this paradox had resulted in the idea that desired behaviour could only be obtained through classic solutions: law enforcement, financial stimuli, communication, social engineering and a good quality of the built environment. Whilst there is no doubt that these measures matter, they were all found within the

confines of the municipal policy toolkit. Instead of truly relying on self-organizing citizens, as was the official wish, these measures boxed citizens into what the officials considered 'desired behaviour'. Desirable behaviour was assumed to lead to responsible citizens and, consequently, to less waste. But the officials overlooked the fact that manufactured desired behaviour does not make one responsible; it merely makes one obedient.

THIRD SESSION AND END OF THE PROJECT

The findings outlined were presented during the third and final session. Contrary to the first two sessions, which had relied heavily on the input of the participants, this session was used by the researchers to present findings. The centrepiece was the final iteration of the model, as developed from the data collected by the officials and discussed during the meetings. Whilst we had done our best to include as many details as possible, it was inevitable that some minor details were lost. Also, some comments made during earlier sessions turned out to be difficult to make sense of.

However, it was not the inclusion or exclusion of details that was noted. The general comments were more targeted at the model as a whole. The participants either found it difficult to follow the many loops in the model or, conversely, thought that the model was too generic to reveal anything surprising. Of the latter group, some remained silent out of politeness but were still deeply sceptical. The details of the model and our observations mentioned in the previous section seemed to fall upon deaf ears. In hindsight, this should not be surprising because these items concern blind spots of the organization's members, for example, searching within the existing problem-and-solution space as dictated by the organization's tasks and subdivisions.

Subsequently, the municipal strategists stressed that they had wanted the researchers to make a model of the urban system that could function as a control panel during budget cuts, a master tool to effectively solve complex societal problems. Note how this differs from the initial

question, which concerned the extent to which budget cuts could be made without sliding into a downward spiral. Attempts were made during the process to temper these expectations because there was virtually no time available and because it is possible to have serious doubts about the feasibility of such a control panel. Because these findings were far removed from the actual—but never voiced—expectations, the project was cancelled.

RESULTS

In this field experiment, we tested to what extent we could compress GMB. We deviated from the basic GMB model in two ways. Firstly, we substituted the researcher-led initial modelling for client-led modelling. We expected that a guided group effort could lead to an equally solid starting point. Secondly, we allowed for a relatively homogeneous group of participants. We expected that it would be possible to instruct the civil servants with regard to data collection and that this would introduce more heterogeneity in the group discussions and, subsequently, in the model. We would assess the output at two levels: the individual level (positive response to the process, mental model refinement, commitment and behavioural change) and the group level (increased quality of communication, creation of a shared language, consensus about problem and solution definitions). Regarding the latter, we did not observe any improvements. The quality of the communication did not advance, and there were no changes to the semiotic gaps in the group. The assessment at the individual level is somewhat different because a small number of participants stated that they could see the utility of systems thinking. The participants who responded positively mentioned the following strengths: becoming clear about causes and consequences in public policy; learning more about waste and cleaning policies, if only because of talking to people they had never talked to before; and engaging in group discussions about policies at a systemic level.

However, we cannot overlook the fact that most participants did not respond positively.

Furthermore, there was a genuine lack of commitment. The respondents that were negative about the method mentioned the following weaknesses: the amount of time and other resources required, the skills demanded from the participants, the resulting model that was either self-evident or not clear enough and lack of a clear goal for using the method. To answer the research question, both modifications can be rejected on the basis of this particular field experiment.

One can argue that the lack of observable changes does not necessarily mean a negative result. However, there are three reasons to evaluate the outcomes as unfavourable. Firstly, the client had aimed for some kind of organizational change. Whilst it is true that the client was unclear about the actual goals of the endeavour, the case can be made that the organization did not achieve any change despite a degree of exposure to the method. Not meeting that goal can be considered an unfavourable result. Secondly, GMB itself aims to provoke change towards system improvement and it is clear that this was not achieved. Thirdly, the client rejected the method after the sessions. Taking these three points together leads us to conclude that the shortcuts did not work. The question then is why they did not work as expected.

The main reason for starting with an expert-developed preliminary model is that the organization had blind spots. We expected the officials to be unaware of those blind spots but believed that they could obtain new ideas and remove blind spots through a careful selection of respondents and good interviewing practices. This proved to be a bridge too far. People who believe that they know all they need to know are unlikely to ask for counterfactuals. This was reflected in their selection of the respondents, most of whom were working in the same organizational unit. For the participants, there was little reason to ask street cleaners or citizens, for example, because, as one participant put it, 'all the knowledge needed is already present in the organization'. In addition, people have a bias towards information that confirms their own views and tend to reject information that

does not fit within that frame. This was particularly strong in this case where the solution seemed to have preceded the root problem definition.

As is more often the case in a public organization, the participants were engaged in a power struggle. Some of them were primarily concerned about the survival of the group or department they represented so they saw the process as an opportunity to push their interests to the fore. It explains why there was a great reluctance to look beyond the borders of their own organizational unit, and it explains why some of the participants were less eager to join the experiment than others. People were more willing to participate if there were a chance that the experiment could further their position. Conversely, people who believed that there was little scope for this pulled out early.

We should also note that there was simply a lack of time. As examples of GMB show, these processes can take considerable time to materialize, although it is not impossible to do it quickly, as Rouwette (2012) demonstrated. However, it cannot be completed in a hurry if the group is inexperienced and insists on doing things themselves so that they can learn despite the lack of time. Most of the participants agreed that they had put little to no effort into the initial data collection and even had little time for joining the group sessions. Besides, political reality dictates that what is important today may not be important tomorrow.

Fourthly, and related to the previous points, the participants may have severely underestimated the efforts required for successful participatory system dynamics modelling. Whilst some of the participants were at least aware of how it should work (on account of having read Senge's, 1990 book), we think that none of the participants appreciated the fact that GMB demands a concerted effort.

Whilst it is only natural that we focus on the performance of the altered method, we should also bear in mind that the organization in question did not provide an ideal testing ground for our ideas. It is necessary to consider the context. The following reasons were offered in a round-table evaluation half a year later. Firstly, it was

noted that there was an organizational fear of trying out new things, which was exacerbated by a constant state of reorganization and a sense of stringent accountability measures. Many plans were hatched, but few were executed because of risk-averse attitudes. This culture also promoted pillarization as employees learned that working within one's own organizational unit yielded lower risks than working across a unit's boundaries. Within one's own department, it is clear what one is accountable for. The urge not to venture across one's own boundaries was reinforced by the fact that new accountability measures did not go hand in hand with a bigger mandate. Mandates were tightened: More and more signatures were needed to account for certain decisions. Subsequently, mutual trust was fading. Mutual trust is a vital condition for bridging perspectives and coming to a shared reality in GMB. This explains why some participants were simply too cynical to care about the results of the process.

Furthermore, institutions that are relatively vulnerable within their environments are more inclined to act on their environments, whilst institutions that are relatively well-insulated from threats in their environments will be more hesitant to do anything about the status quo. Municipal authorities have few natural threats to their survival as they have no natural competitors, apart from some political strife. Therefore, they are less inclined to act on their environment. They will survive, even if their impact on that environment is far from perfect. This may help to explain why the project was unsuccessful.

It may be unusual to report on a failed GMB project, but it is instructive to see under which conditions a systemic approach to policy analysis in general, and the deployment of GMB in particular, can fail. Even an ostensibly simple policy issue is driven by a multitude of interlocking factors that can be easy to become entangled in. The delicate participative method requires time and energy to uncover those factors, and both were in short supply. We may have been naïve in stepping in and trying it regardless, but such experiments are necessary to further our understanding of GMB.

ACKNOWLEDGEMENT

The authors would like to thank Roy Heidelberg (Louisiana State University) for reviewing an earlier version of this article.

REFERENCES

- Andersen DF, Richardson GP. 1997. Scripts for group model building. *System Dynamics Review* **13**(2): 107–129.
- Andersen DF, Richardson GP, Vennix JAM. 1997. Group model building: adding more science to the craft. *System Dynamics Review* **13**(2): 187–201.
- Andersen DF, Vennix JAM, Richardson GP, Rouwette EAJA. 2007. Group Model Building: problem structuring, policy simulation and decision support. *Journal of the Operational Research Society* **58**(5): 691–694.
- Bérard C. 2010. Group Model Building using system dynamics: an analysis of methodological frameworks. *The Electronic Journal of Business Research Methods* **8**(1): 35–45.
- Byrne D. 2002. *Interpreting Quantitative Data*. Sage: London.
- Byrne D, Callaghan G. 2014. *Complexity Theory and the Social Sciences*. The state of the art. Routledge: Abingdon.
- Cavana RY, Boyd DM, Taylor RJ. 2007. A systems thinking study of retention and recruitment issues for the New Zealand Army electronic technician trade group. *Systems Research and Behavioral Science* **24**(2): 201–216.
- Checkland P. 1981. *Systems Thinking, Systems Practice*. John Wiley: Chichester.
- Checkland P. 2011. Autobiographical retrospectives: learning your way to 'action to improve' — the development of soft systems thinking and soft systems methodology. *International Journal of General Systems* **40**(05): 487–512.
- Checkland P, Howell S. 1998. Action research: its nature and validity. *Systemic Practice and Action Research* **11**(1): 9–21.
- Delbecq AL, Van de AH V, Gustafson DH. 1975. *Group Techniques for Program Planning: A Guide to Nominal Group and Delphi Processes*. Glenview: Scott, Foresman and Co.
- Flood RL. 1999a. *Rethinking the Fifth discipline: Learning Within the Unknowable*. Routledge: London.
- Flood RL. 1999b. Knowing of the unknowable. *Systemic Practice and Action Research* **12**(3): 247–256.
- Gerrits L. 2012. *Punching Clouds. An Introduction to the Complexity of Public Decision-Making*. Emergent Publications: Litchfield Park, AZ.
- Hovmand PS, Andersen DF, Rouwette E, Richardson GP, Rux K, Calhoun A. 2012. Group Model-

- Building 'scripts' as a collaborative planning tool. *Systems Research and Behavioral Science* **29**(2): 179–193.
- Klijn EH, Snellen ITM. 2009. Complexity theory and public administration: a critical appraisal. In *Managing Complex Governance Systems*, Teisman GR, van Buuren MW, Gerrits LM (eds.). London: Routledge; 17–36.
- Richardson GP, Andersen DF. 1995. Teamwork in group model building. *System Dynamics Review* **11**(2): 113–137.
- Richardson KA, Midgley G. 2007. The evolution of systems thinking. *Systems Theory and Complexity, part 4* **9**(1–2): 163–180.
- Rouwette EAJA. 2012. Does Group Model Building work? Evidence from and comments on the paper by Videira et al. *Systems Research and Behavioral Science* **29**(6): 620–623.
- Rouwette EAJA, Vennix JAM. 2006. System dynamics and organizational interventions. *Systems Research and Behavioral Science* **23**(4): 451–466.
- Scott RJ, Cavana RY, Cameron D. 2015. Mechanisms for understanding mental model change in Group Model Building. *Systems Research and Behavioral Science*. doi: 10.1002/sres.2303 (early view).
- Senge PM. 1990. *The Fifth Discipline: The Art and Practise of The Learning Organization*. Doubleday: New York.
- Tsoukas H, Hatch JM. 2001. Complex thinking, complex practice: the case for a narrative approach to organizational complexity. *Human Relations* **54**(8): 979–1013.
- Turner BS. 2005. Talcott Parsons's sociology of religion and the expressive revolution. *Journal of Classical Sociology* **5**(3): 303–318.
- Vaandrager D, Gerrits L, Bressers N. 2014. Reconstructing the Blue Connection: a systemic inquiry into the improbable project of building a new river through a (sub)urban area. *Systems Research and Behavioral Science*. doi: 10.1002/sres.2297 (earlyview).
- Van Nistelrooij LPJ, Rouwette EAJA, Verstijnen IM, Vennix JAM. 2015. The eye of the beholder: a case example of changing clients' perspectives through involvement in the model validation process. *Systems Research and Behavioral Science* **32**(4): 437–449.
- Vennix JAM. 1996. *Group Model Building, Facilitating Team Learning Using System Dynamics*. John Wiley & Sons: Chichester.
- Vennix JAM, Andersen DF, Richardson GP, Rohrbaugh JW. 1992. Model-building for group decision support. Issues, and alternatives in knowledge elicitation. *European Journal of Operations Research* **21**(4): 331–347.
- Wagenaar H. 2007. Governance, complexity, and democratic participation: how citizens and public officials harness the complexities of neighborhood decline. *The American Review of Public Administration* **37**(1): 17–50.