New Mathematics and Natural Computation
Vol. 7, No. 3 (2011) 515–542

DOI: 10.1142/S1793005711002050

VALIDATION AND AGENT-BASED MODELING: A PRACTICE OF CONTRASTING SIMULATION RESULTS WITH EMPIRICAL DATA

FRANK C. S. LIU

Institute of Political Science, National Sun Yat-Sen Univeristy
No. 70 Lian-hai Road, Kaohsiung City 804, Taiwan, ROC
csliu@mail.nsysu.edu.tw

As an emerging approach to explore the dynamics of voter preference, agent-based modeling (ABM) highlights new opportunities for intellectual exchange across disciplines, such as mathematics, political science, communication studies, and computer science. By aiming to contribute to cross-disciplinary communication for a better application of this approach, this paper summarizes what scholars have done about internal and external validation and presents a comparison between statistical analysis based on datasets generated in a laboratory and analysis based on corresponding empirical datasets. The results of the comparison suggest that, although there is no perfect matching, the comparison reveals some similarities in terms of increase or decrease in the proportion of different types of agents. This result further implies that an internally valid ABM model may lead to a certain level of external validity.

Keywords: Agent-based modeling; empirical validation.

1. Introduction

Agent-based modeling (ABM) as a methodology for studying complexity has been applied to ecology, physics, biology, economics, management, anthropology, engineering, sociology, and psychology over the past few decades. It has also attracted attention from political scientists and resulted in creative applications in this discipline in the late 1990s.[1–8] Having originated in the Santa Fe Institute through intelligent debates between economists and physicists in the late 1980s, the basic idea behind ABM is to use an object-oriented computer language, such as C++, Objective-C, or JAVA, to create self-organizing objects (i.e. a set of commands and variables that can take actions, process information, and make decisions based on given rules). The purpose in creating an agent-based model is to observe visual or statistical patterns at the aggregate level emerging from automatically interactive agents.[9–11]

Due to its computational and experimental nature, ABM is widely welcomed by the disciplines that emphasize objective analysis of the external physical world, such as ecology and biology. In contrast to field experiments, where researchers use convenient samples to observe the effect of certain treatment operations on subjects in naturally-occurring environments, ABM is labeled as a laboratory experiment, and one that is completely based on computing. Despite this difference, both lab and field experiments emphasize the randomization of subjects (or other sampling units) and the comparison of outcomes across groups under different circumstances. Therefore, the discussion on field experiments usually sheds light on lab experiments.

A general concern shared by social scientists and humanities scholars is the applicability and predictability of experiments to the empirical world. Empirical validation has been an issue for scholars of psychology, management, and experimental economics, particularly those who attempt to apply lab experiment results to the real world.[12,13] For political scientists, the discussion on ABM has concerned how it is used to aid intuition or facilitate a "thought experiment".[14] In his *The Complexity of Cooperation: Agent-Based Models of Competition*, Robert Axelrod suggests that ABM "does not aim to provide an accurate representation of a particular empirical application. Instead, the goal … is to enrich our understanding of fundamental processes that may appear in a variety of applications".[14,p.5]

Over the past few decades, most social scientists using ABM have been following this logic of thought experiment because of (1) the lack of empirical data for contrasting empirical findings with the findings derived from experiments, and (2) a greater interest in solving theoretical puzzles than empirical puzzles. Applications of ABM in the social sciences, therefore, are commonly centered around issues that do not require a serious empirical validation process, at least compared to empirical data, such as theoretical circumstances where diffusion of identity would occur,[15] where political disagreement would survive,[16] and where regimes would survive,[8] etc.

Even though these pioneering works have generated greater interest in applying ABM to examining existing theories, one needs to notice that there is a group of scholars encompassing different disciplines paying greater attention to a more fundamental question regarding this approach: Can my simulation results be communicable to those who distrust the results of my experiments and laboratory work? Specifically, to what extent can the results of such laboratory experiments be associated with the empirical data, if they are available?[12,17–24] As Schram concludes in his observation, "a search for ways to modify laboratory experiments seems a fertile research area now that experiments are starting to be used for much more than just testing theories".[12,p.235] ABM researchers have started to look for ways to conduct their research out of laboratories.

This paper, as a way of contributing cross-disciplinary communication about methodology, attempts to address this "how to apply ABM correctly if external validation is concerned" question by (1) summarizing works from different disciplines and finding a commonly agreeable procedure to conduct validation of an ABM experiment, and (2) providing an example of applying some elements of this procedure to public preference research.

After an overview of the literature regarding validation and the methods used to increase the validity of an ABM project, this paper details how an agent-based model is aligned to John Zaller's Receive-Accept-Sample (RAS) theoretical framework of preference formation.[25] Next, it will show how this model is calibrated with the first wave of survey data collected during an electoral campaign. The aggregate level statistics and patterns derived from the simulation results will then be contrasted with descriptive statistics drawn from the data for the second-wave survey conducted after the election. The final section will discuss the lessons learned from the comparison.

2. Validation of ABM: Internal and External Validity

In general, the term "validity" means a good correspondence between a real system and an artificial model. Laurent suggests that a clear understanding and description (or "the verbal model") of the real system through qualitative methods (e.g. interviews) is necessary before designing a mathematical, experimental, or computational model like ABM.[20] By a rigid definition, a validated ABM model should be both internally and externally valid. An internally valid model is the one from which researchers can draw confident causal conclusions, and so such a design will yield robust and replicable results. Mckelvey refers to internal validity as "analytical adequacy," meaning that "the model (in an isolated idealized setting, such as a lab or computer) correctly produces effects predicted by the theory".[26,p.766] A different but more relaxing definition of internal validity is having a consensus about the model design from members of the associated academic community.[27,28] As long as the behavior rules and environmental settings of a model make sense, this model can be seen as being internally valid.[29–31]

An externally valid model allows a researcher to generalize conclusions to situations that prompted the research, including from a sample to a larger populations, generalizing across settings or population, and both.[12,21] In particular, when an assumption held by a model is supported by robust empirical evidence, this model will have a higher degree of methodological realism.[32] McKelvey refers external validity to "ontological adequacy," meaning that a model passes the process of testing the model against evidence from the real world.[26]

In effect, however, there exists no model that is constructed on the basis of complete information or the whole picture of a story. Details of a case told by field researchers may still not be complete enough to form an agent-based model; similarly, a model that is designed on the basis of empirical findings may still have many unspecified assumptions when it comes to realization in ABM. Laurent hence argues that ABM modelers should not be worried about not being able to tell the whole story.[20]

Laurent points out that tension exists between internal validity and external validity, which means that a good internally valid model may sacrifice its external validity. "Where internal validity often requires abstraction and simplification to make the research more tractable, these concessions are made at the cost of decreasing external validity".[12,p.226] Schram points out that such conflict and artificiality exist in four areas or goals of experimental research.[12] From low to high demand of external validity, these four areas of research are (1) testing theory, (2) theory stress tests, (3) searching for empirical regularities, and (4) advising policy-makers. As mentioned above, the goal of testing theory (area No. 1) requests the least external validity of an ABM model, while the goal of advising policy makers (area No. 4) demands the most external validity.

Theory stress tests (area No. 2) are usually used in experimental economics, where experiments are used to test how institutions work and whether the behavioral assumptions underlying a theory hold. An agent-based model is used to test extreme situations (i.e. by using extreme values for chosen parameters) while relaxing some of the assumptions. This type of research allows researchers to explore the domain of applicability of a thought or a theory. Hence, to scholars conducting theory stress tests, a theory that holds in extreme situations will gain external validity. However, the major problem of this type of experiment, as with contests, is increase in artificiality because this approach reduces the need to focus on the external validity.

As regards to searching for empirical regularities (area No. 3), experimental economists use experiments to find empirical regularities in the field where no theory has yet formed or a theory has been rejected, such as observed phenomena (e.g. groups of individuals are better at solving optimization problems than individuals are) or observed causal effects (e.g. training in economics leads to more selfish behavior). Apparently, this type of research requires more emphasis on external validity than the previous two types and needs to deal with more critics about external validity. As Schram warns, "If the aim is to make any claims about other regarding preferences in the world at large, the high artificiality appears to render the experimental results useless. If, in contrast, the aim is to document robust, causal laboratory effects to confront theories, artificiality is less of a problem".[12,p.233] All this implies that a modeler aiming to find empirical regularities with experiments needs guidance from the empirical world that he or she is trying to study.

When there is no theory directly addressing empirical or policy concerns, experiments are employed to provide some suggestions (area No. 4). The major problem with offering advice based on simulation results is that "it is very difficult to judge *a priori* whether or not the external validity is high enough".[12,p.234]

For Schram and Laurent, theory plays a less important role if the goal of the lab experiment is searching for empirical regularities (area No. 3) or advancing policy-makers (area No. 4). However, one often finds disagreement about whether a theory should be connected to an experiment in these two types. McQuarrie, for example, emphasizes "theoretical labeling" in the design of an (field) experiment, that is, it is necessary for modelers to link external validity to construct validity (the extent to which measures accurately reflect the theoretical concepts they are intended to measure).[22] This argument implies that theory should be considered in all of the four areas of the experiment, even though the goals of research may differ. "A belief in validity trade-offs allows an investigator to affirm the importance of any given type of validity while continuing to ignore it in practice".[22,p.143] Moreover, he contests that modelers should consider external validation at the outset of designing a model:

> "By assuming this separation, theory research is freed from having to worry about concrete particulars, and issues of proximal similarity become pertinent only within the subsequent stages of intervention and effects research, where generalizing to a narrow subset of cases is the focus .... How we label our treatment operations−the stuff of construct validity−determines which subsets of other treatment operations we have warrant to generalize to and cross−the stuff of external validity. Conversely, our successes and failures with respect to achieving external validity help us decide whether our initial labeling of the treatment operation was correct. Hence, under the assumption that treatment operations reflect multiple constructs, laboratory tests of theory must concern themselves with external validity and will naturally do so as part of vigorous efforts to secure construct validity".[22,p.148]

Looking from a logical positivist perspective, Lucas also argues that the external validity of experiments should be rooted in theory instead of methodological procedures.

> "The major drawback to experimental investigations is that, at times, they cannot create or manipulate all theoretically meaningful variables. However, if an experiment does manipulate every theoretically relevant variable and finds an effect, then to say that the effect will not generalize to naturally occurring situations is not a criticism of the experiment as having low external validity; rather, it is a critique of the theory for not taking every factor influencing the phenomenon of interest into account".[21,p.238,a]

Despite the discrepancy, one can see that "alignment with theory" is a shared commonality between the two conflicting perspectives regarding the role of theory in experiments. Laurent, in emphasizing the connection to the empirical world, suggests that a modeler describe (1) the characteristics of the real life system, such as retailing costs in economic analysis, (2) the assumptions, and (3) empirical or factual information that "re-formalize the model in a more realistic manner," such as real merchandising costs in real stores.[20,p. 181] Lucas, in emphasizing the role of theory, thinks that well-designed experiments (1) simplify naturally occurring situations and (2) incorporate only theoretically relevant elements. The key to evaluating any research

[a] Although Lucas emphasizes the importance of theory, he holds that there is a trade-off of external validity — "the more particular information we gather about a specific phenomenon, the less that information will generalize to other phenomena".[21,p.246]

design is "whether the situation adequately represents the concepts or relations specified in the theoretical hypotheses under test".[21,p.246,b]

To sum up, an important step into constructing the validity of the model, which reflects a newly-formed consensus among scholars using lab experiments, is to evaluate carefully the theoretical elements of a simulation model. In particular, when the goal of research is not making an empirical prediction, or when data for external validation are unavailable, researchers are expected to carefully choose parameters derived from theory (construct validity) and, if the theory does not explicate some assumptions or parameters, then formulate assumptions based on empirical findings (analytical adequacy).

3. Research Methodology for Conducting External Validation

Beyond the alignment of theory, Methta and Bhattacharyya suggest that researchers need to consider two more stages to fulfill the requirements of external validation: the alignment of observable processes and the alignment of outputs.[23] Above how to deal with these two stages correctly, experimental economists have formed three approaches[33]: the indirect calibration approach (IC), the Werker-Brenner approach (WB), and the history-friendly approach (HF).[c]

The indirect calibration (IC) approach includes four steps:

- Step 1. The researcher identifies a set of stylized facts that he/she is interested in reproducing and/or explaining with a model.
- Step 2. The researcher builds the AB model in a way that keeps the description about agents and rules as close as possible to empirical and experimental evidence regarding the agents and rules. That is, the design of the model should gain support either indirectly from the literature of the subject of interest, or directly from the empirical data collected.
- Step 3. The researcher uses empirical evidence regarding stylized facts to restrict the space of parameters and determine whether the statistical regularities derived from simulation are consistent with the empirically-based stylized facts of interest.[d]

[b] Here are more details about the criteria he suggests to access external validity: (1) construct validity: the extent to which measures accurately reflect the theoretical concepts they are intended to measure; (2) relevance: the degree to which the situation designed to test the theory adheres to the scope of conditions of the theory being tested; (3) reproducibility: whether the study can be repeated in the same conditions and still produce the same findings; (4) consistency: the extent to which the observations in a study are consistent with each other and with the theory being tested. (i.e., whether the findings of the study support the theoretical proposition(s) being tested); and (5) confirmatory status: the extent to which the proposition has been supported by numerous tests in diverse settings.

[c] Note that the steps summarized below are not supposed to be fixed standards or rigid guidance to validate an experiment; instead, they together show how scholars across disciplines think about the procedure of the experiment.

[d] In this step of reducing parameters, "one must generate a distribution for the statistics summarizing the stylized facts of interest … and test the null hypothesis that the empirically observed values can be generated by our model under that particular parameter combination".[33,p.23]

- Step 4. The researcher turns to find statistical regularities or patterns that contradict, or at least features that are different from the stylized facts characterized by the empirical data.

The Werker-Brenner approach (WB) includes three steps:

- Step 1. The researcher uses empirical data or empirical knowledge to narrow down conditions, to reduce dimensions, or to calibrate initial conditions and the ranges of the model parameters. In this step, a number of parameters are chosen for the next stage of validation.
- Step 2. The researcher uses the model specification and generates a Monte Carlo set of micro and macro time series data for that particular combination of empirically plausible parameter values. This will result in a set of theoretical realizations for each model specification. This is the step parallel to the sensitivity test to be described below.
- Step 3. The researcher compares between theoretical realizations and empirical realizations (i.e. real-world data). Scholars using this approach advocate the use of Bayesian inference procedures to validate the outcomes. "Empirically observed realizations are used to further restrict the initial set of model specifications (parameter values) that are to be considered. The modeler only retains those parameter values (i.e. model specifications) that are associated with the highest likelihood by the currently known facts (i.e. empirical realizations). Model specifications that conflict with current data are discounted" (p. 24).

The history-friendly approach (HF) includes three steps:

- Step 1. The researcher conducts specific historical case studies about the macro-level subject (e.g. an industry) to model parameters, agent interactions, and agent decision rules.
- Step 2. The researcher conducts sensitivity analysis to examine if "history divergent" results are possible.
- Step 3. The researcher compares the output (the "simulated trace history") with the actual history of the industry.

Concerning the availability of long-term data with rich information about the parameters of interest, IC is preferable to WB and HF for the social sciences. Both WB and HF emphasize narrowing down the set of parameters with empirical data. As a result, the calibration process of both approaches is restricted by data availability and reliability. Datasets used in political science research, however, are usually limited in scope and duration. The number of parameters required to design an agent-based model usually exceeds what survey research can provide. Particularly in public opinion research, the choice of variables is strictly limited by the questionnaire's length and the questions asked. Even though HF justifies the external validity of the simulation results better than the other approaches, it requires an abundant amount of empirical data and only applies to macro-level inquiries.

Moreover, the way in which researchers use WB and HF to fit parameter values to empirical data (Steps 2 and 3) gives rise to a questionable assumption: data collected for external validation are unbiased or generalizable. Apparently, this is not usually the case in survey datasets. It is difficult to use survey datasets to determine the right time point to start a simulation and a right time point to stop the process. In addition, even given proper data, a modeler will still not have a clue about the underlying connection between the data set and simulation results.[e]

From the above discussion, one can see that IC is a comparatively proper approach for ABM modelers in political science. First, IC better fits the requirement of "theory alignment," a goal emphasized in this discipline. WB and HF require researchers to calibrate parameter values before conducting validation, while IC requires that (internal, theory-oriented) validation precede calibration.[33] Second, IC is more flexible for political scientists to use because it is not a requirement that political scientists use empirical data to calibrate the parameters of a model.

Such flexibility is based on two understandings. First, researchers know that, even if we restrict the parameter space, i.e. limit the number of parameters or variables used in the model, the details of the model can hardly be compared with empirically-observed ones. Second, even if we obtain a suitable sub-region of parameter values that are able to replicate the set of stylized facts of interest (Step 3), it is still difficult to interpret "all comparative exercises that aim at understanding what happens when one tunes the parameters within that sub-region".[33,p.23] In particular, it has been said that "alternative parameter values in an evolutionary world where history, indeterminacy, and non-linear feedbacks between the micro and macro levels may strongly affect the outcomes".[33,p.24] Hence, the IC approach may not help an ABM modeler to achieve a level of external validity as high as that with the WB and HF approaches because of its ignorance regarding incorporating empirical cases or history into model design.

The above approaches provided by experimental economists correspond to those provided by scholars in marketing science. The four steps provided by Garcia *et al.*[18] and Methta and Bhattacharyya[23] include:

- Step 1. Grounding, including delineating the model's scope based on qualitative and quantitative data (face validity), assigning realistic characteristics and initial conditions to an agent (parameter validity), and making sense of the overall model simulation on a macro-level (process validity).
- Step 2. Calibration: fine-tuning the simulated model to some particular unrelated features of historical data that are drawn from macro-level data.
- Step 3. Verification: checking graphically or statistically to see if the results of the simulation capture the intent of the real-world model, that is, to see if the results are consistent with the stylized pattern shown in the empirical data at both the individual level and aggregate level.

[e] See Ref. 33, pp. 25–27 for more discussion.

- Step 4. Harmonization: if the goal of an experiment is to forecast, researchers need to compare the predictive outputs of the commutative model with the predictions of a statistic model.

To sum up, the key elements drawn from the above approaches include: (1) describing the choice of a limited number of parameters based on a theory, (2) constructing the correspondence between these elements of the model to a targeted theory, (3) employing empirical data to calibrate initial conditions and parameters, (4) evaluating how robust and sensitive the values of key parameters are, and (5) discovering how much the statistical patterns of simulation correspond to those found in empirical data.

3.1. *Comparison across patterns*

The above overview summarizes what scholars from other disciplines have suggested to do with the alignment of theory and the alignment of observable processes. According to the IC approach, after aligning the model with a theory and carefully calibrating parameter values, the last step is "aligning outputs" or contrasting simulation results with empirical findings.

A common way of drawing comparisons in recent ABM studies involves presenting a social phenomenon that corresponds to the patterns found during the simulation. Mpwal *et al.*, for example, in their study on the emergence of opinion clusters, employ a variety of empirical evidence to show the similarities between the founded opinion clusters and social clustering seen in the elections in Germany and the U.S. elections.[34] This helps a reader to quickly makes sense of the lab reports. The shortcoming of such a comparison, however, is that the author(s) will be free from the burden of explaining what went on in the formation of the patterns and corresponding empirical phenomena. As they acknowledge, "although there are clear empirical referents for polarization, it is less clear where to turn for real-world evidence of clustering. Perhaps clustering is so pervasive that we often fail to notice it".[34,p.369]

As Janssen and Ostrom suggest, there are four ways in which simulation results can be compared with stylized facts.[19] First, if the data are abundant in quantity and good in quality, and if one can draw stylized facts from such empirical data, researchers can derive statistical distributions, such as the power law distribution, and other stylized facts from the empirical data.[2] Second, if the model is relatively uncomplicated, researchers could ask what the simple rules in the model that generate these stylized facts are. After finding out these rules, the next step is to investigate the modeled conditions that result in statistics similar to the observed styled facts. Third, researchers can use the empirical findings from role games or field experiments to develop and test assumptions held in ABM. Fourth, the researchers can case studies with rich information and data to parametrize the model (conjoint analysis).[18]

Besides comparing stylized facts or patterns, sensitivity analysis or robustness tests (based on the IC approach) is an important step that makes an ABM experiment communicable across disciplines. Robustness tests make the findings more persuasive to readers who are concerned with the influence of extreme conditions during simulation. Because parameter values are used to test whether the simulation results are stable in extreme scenarios, they open the modelers "to the propositions that a model should be judged by the criteria that are used in mathematics: i.e. precision, importance, soundness and generality".[33,p.22]

Note that, as Guala warns, a robustness test is not an external validation for mixing robustness tests with external validation can cause confusion.

> "By adding details to the most basic theoretical models, one in a way proceeds toward a level of analysis that is more 'concrete' and closer to application. But such 'middle-range' models can rarely be applied directly to the functioning of a specific economic system (i.e., unless they are further modified to take into account more context-specific factors).... The main difference has to do with the absence/presence of a concrete specific target system: whereas external validity requires the identification of such a target, robustness arguments do not".[35,p.225]

Robustness tests provide a range of possible situations for empirical evidence to fit, but whether or not the model is externally valid still depends on how much the selected criteria of patterns, and the mechanisms causing the patterns, match an empirical case, and not on the robustness tests themselves. "External validity inferences do not have much bite unless one systematically investigates the degree of similarity and dissimilarity between laboratory and target systems".[35,p.229]

Given all the above discussion drawn from the literature, the second half of this paper will present an example of applying the IC approach to a study of preference dynamics. As each approach has its limits and because no agent-based model is perfect at the time created, the example should not be regarded as a perfect one. Instead, through this practice, one will more clearly see the limits of ABM and what needs to be done to advance an agent-based model and to validate the simulation results.

4. Alignment of Theory

The Swarm-RAS model, or S-RAS, is a "Swarm" version of John Zaller's Receive-Acceptance-Sample (RAS) theory of voter preference.[25,f] It has been difficult to study the dynamics of preference formation with RAS where the axioms of information processing should be considered together. I designed S-RAS with a view to reviving this framework that has been more cited than directly applied.

S-RAS allows researchers to operationalize the original RAS theory. Using S-RAS, multiple agents that are able to process information in R-A-S fashion will be put into

[f] Swarm is one of a number of promising and free tool-kits for ABM. It is available on line at www.swarm.org.

an artificial society where they will interact with self-selected news media, political experts, and other fellow citizen agents. Moreover, S-RAS considers real-world complexity like individual differences by including three types of individuals — the politically aware, the politically unaware but with clear electoral preferences, and the politically unaware with no preferences regarding candidates.

Given these elements, hopefully, a researcher can use S-RAS to study the long-term influence of a specific element of the model. The next subsection describes how the model design of S-RAS was associated with the RAS framework, particularly the four basic axioms of information processing.

4.1. *The rules of information processing*

John Zaller's RAS framework provides clear rules regarding how individuals process political information. It provides a solid base for constructing an ABM model of voter preference. RAS is composed of four axioms of information processing: reception, resistance, accessibility, and sampling. The reception axiom (Axiom 1) states that political experts, including political elites, are aware of political issues and are more likely to acquire political information actively. The level of their political awareness determines the probability of obtaining political information.

The resistance axiom (Axiom 2) indicates that this awareness of political information determines the propensity to reject incoming political information. Given an assumption that individuals relate an issue to their political predisposition, this axiom suggests that the more politically aware they are, the more likely they will resist incoming political information. Hence, political experts who are better-informed tend to resist information they encounter, whereas the majority of voters who are poorly informed tend to accept whatever information they encounter.

The last two axioms account for how individuals access obtained information and form preferences. The accessibility axiom (Axiom 3) attests that individuals recall messages off the top of their head or base their statement of preference on information recently stored in their memory. The response axiom (Axiom 4) testifies that individuals sample the stored messages to form their attitudes by "averaging across the considerations that are immediately salient or accessible to them".[25,p.49]

The above axioms suggest that individuals follow the Receive (Axiom 1)-Accept (Axiom 3)-Sample (Axiom 4) procedure to form their opinions about a political issue or a candidate. Theoretically, it remains unclear if individuals resist information before or after they receive it. The resistance axiom, hence, is incorporated in the model design in a less explicit fashion than the other two axioms. To validate Axioms 1 and 2, Zaller examines empirical data sets and finds that the reception of political information is a function of political awareness and that the resistance of political information is a joint function of political awareness and predispositions. Unfortunately, this still proves that individuals process information in the order of Axioms 1 and 2. While political awareness refers to political knowledge, "predispositions are the critical intervening variable between the communications people encounter in the mass media, on one side, and their statements of political preferences, on the other".[25,p.23,g]

4.2. *The operationalization of key variables in RAS*

One important task of ABM is transforming theoretical concepts into behavioral rules and features of agents. In S-RAS, seven parameters (or variables) are necessary to make RAS functional in an ABM environment: *Partisanship, Voter Preference, Opinion, Political Expertise*, *Propensity to Access the News Media*, *Propensity to Discuss Politics*, and *Propensity to Perform Selective Perception.*

An agent's *Partisanship* by design is immutable, while *Voter Preference*, as a function of the moving average of *Opinion*, presumably varies. (The relationship between *Opinion* and *Voter Preference* is critical to model design and will be fully discussed in the next subsection.) *Partisanship* in S-RAS corresponds to the variable "political predisposition" in RAS. An agent's *Partisanship*, denoted by 0 or 1, refers to his identity with Party 0 or Party 1 in a two-party system. Agents of *Partisanship* "0" will be initiated with *Opinion* favoring the candidate of Party 0 (and, of course, will form *Voter Preference* for "0"). This agent will also like to find discussants with "0" identification, and when performing selective perception (i.e. resisting incoming information from the media or reinterpreting information from the news media to be consistent with their *Partisanship*)[37] will (re)interpret and save received messages as "0". All of these rules apply to those with partisanship "1".

The parameter *Political Expertise* refers to an agent's political knowledge, with the concept being positively related to the level of political awareness in Axiom 1.[h] In S-RAS, every agent has a randomly assigned value of *Political Expertise*, ranging from 1 to 10. Political experts, or the politically aware, are higher in terms of this parameter value than the politically unaware.[i] In S-RAS it is assumed that those with

[g] This last axiom is based on a model in psychology, which emphasizes that the current impression about one thing is an average of stored impressions. The alternative model is online information processing, or the online-based model. In the online-based model, people use a "judgment operator" to continuously update their attitudes as they acquire new information: "They store the updated attitudes in memory and retrieve them in a given situation, including interview situations".[25,p.50] Zaller admits that neither the memory-based model nor the online model describes all cases; instead, it depends on the level of the issue and the availability of information in their memory.[25,p.279] For more discussion on the difference between the memory-based model and the online model, see Ref. 36. The S-RAS model bridges this caveat by adopting the auto-regressive influence model.

[h] Political awareness refers to "the extent to which an individual pays attention to politics and understands what he or she has encountered". It "denotes intellectual or cognitive engagement with public affairs as against emotional or affective engagement or no engagement at all" (p. 21). The measurement of political awareness Zaller chooses is political expertise, or simple neutral or factual knowledge about politics (among other concepts such as cognitive complexity, political involvement, attentiveness, sophistication, and political acuity).

[i] The conventional wisdom in political science holds that the level of an individual's political knowledge is usually stable over time. It is expected that the knowledge gap between information haves and have-nots will not be narrowed in the long run.[38,39] Hence, although it is arguable whether the number of political experts will increase during a campaign season, an assumption held in S-RAS is that the level of an agent's political expertise remains stable throughout a simulation period.

higher values in *Political Expertise* have higher values in *Propensity to Access the News Media*, *Propensity to Discuss Politics*, and *Propensity to Perform Selective Perception*. (These propensities are denoted by values between 0 and 1.) In other words, the politically aware are likely to access the news media, discuss politics with others, and are more likely to reinterpret information as being congruent with partisanship.

4.3. *Opinion formation in S-RAS*

Figure 1 summarizes how agents in S-RAS acquire information and process received messages. In each *time step* agents complete a loop of action, from whether or not to access a news media object to update its voter preference.

As Fig. 1 shows, the following features of agents in S-RAS are initiated: a party identification (1 or 0), an opinion about candidates (0.00 to 1.00), a voter preference (1 or 0), a favorite media object (1 or 0), and eight political discussants with political experts on the top of the agent's contact list. Note that it is assumed that agents habitually check out news media reports before finding someone to discuss politics.[40] For every time step or iteration during the simulation, every agent finishes its own loop of processing political information, including accessing the news media, discussing politics, or doing nothing if no discussant is available.

It is assumed that in S-RAS, agents are concerned with one issue during an election season, such as a choice between Bush and Kerry in the 2004 American presidential election. Specifically, every agent has an *Opinion*, a continuous variable with value varying between 0.00 and 1.00, drawn from a normal distribution. The value of *Opinion* of agents who favor "0" is drawn from a normal distribution between 0 and 0.5; the value of *Opinion* of agents who favor "1" is drawn from a normal distribution between 0.5 and 1.0; and the value of *Opinion* of agents who have no preference in the first place is set to 0.5.

An agent's current opinion, as Axioms 3 and 4 in RAS suggest, is a moving average of voter preferences (1 or 0) perceived from other agents, political experts, and the news media. Axiom 4 suggests that an agent's current opinion can be written as a moving average function, calculated by $\frac{D}{C+D}$, where D denotes the number of dominant messages (i.e. messages that are more intense during the period of attitude change) and C denotes the number of countervailing messages (i.e. the less intense messages). The denominator $(C + D)$ is referred to as the *Capacity to Store Messages* or memory capacity in S-RAS. If an agent X's capacity to store (or remember) messages is 10 and the impressions it has received from other agents or the news media are (1, 1, 0, 0, 1, 0, 1, 1, 0, 1), X's opinion value of the current time step ($t1$) is 0.6. If agent X finds a political expert Y and perceives that Y favors "0," then X's opinion of the next time step ($t1$) will be 0.5 and its *Voter Preference* has changed from 1 to 0.5.

In sum, the value of *Opinion* can be seen as a "true" preference of an agent about a candidate, which other people hardly catch accurately, while *Voter Preference* is

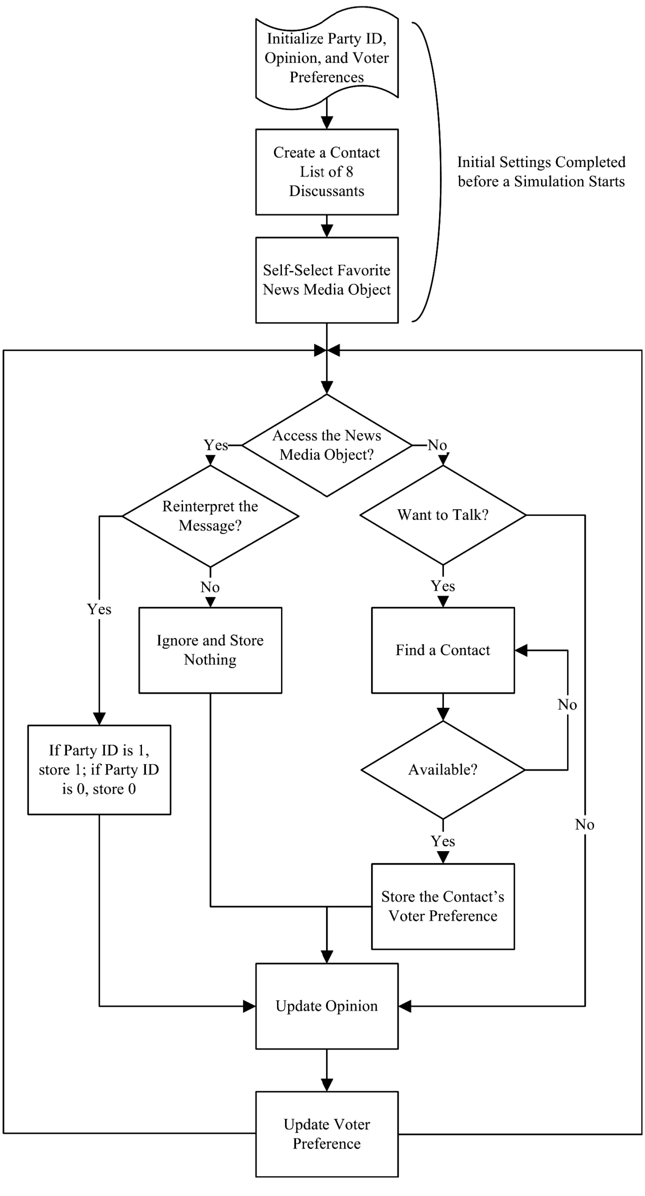

Fig. 1. The flow chart of R-A-S information processing.

better viewed as a general impression of one's true preference that others can tell from dyadic interaction. If agent X has *Opinion* value 0.62, for example, other agents who interact with X will obtain an impression that agent X favors "1". Similarly, if it has *Opinion* value 0.37, its network members who interact with it at a given time will store 0 in their memory.

A change in *Voter Preference* is a consequence of a change in *Opinion*. Suppose agent X favors 1; if its *Opinion* falls below 0.5, its *Voter Preference* will be 0; if X's *Opinion* goes above 0.5, its *Voter Preference* will be 1.[j]

4.4. *Individual differences*

S-RAS maximizes individual differences among agents by considering a general distinction between political experts and ordinary citizens[38,42–44] and interpersonal differences within each type.[37]

As Table 1 summarizes, there are three types of agents in S-RAS and researchers can vary the proportion of each type of agent for the calibration and robustness tests: ordinary citizens with clear voter preference ($C1$ agents), political experts or the politically aware ($C2$ agents), and ordinary citizens having no preference about candidates or claiming independence ($C3$ agents). It is assumed that all features of $C3$ agents are the same of those of $C1$ agents (in computer language, $C3$ agents inherit $C1$ agents), except that the opinions of $C3$ agents are initiated with 0.5 and their party identification is assigned randomly.

Individual differences are characterized by six dimensions: the level of political expertise, the propensities to access the news media, to discuss politics, to be selective about news messages, and the capacity to store messages, and initial opinions. Compared to its $C1$ and $C3$ counterparts, a $C2$ agent has a higher level of political expertise. It is thus more likely to access the news media, to discuss politics, and to perform selective perception. In addition, a political expert has greater capacity to store messages and it bases its opinion on its most recent 20 impressions ("1"s or "0"s) stored in its memory, while an ordinary citizen's opinion is based upon the most recent 10 impressions it has collected.

Table 1. The differences between the three types of agents.

	Ordinary Citizens ($C1$ Agents)	Political Experts ($C2$ Agents)	The Independent ($C3$ Agents)
Political Expertise	[1,5]	[6, 10]	[1,5]
Propensity to Access Media	[0.1, 0.5]	[0.6, 0.9]	[0.1, 0.5]
Propensity to Discuss Politics	[0.1, 0.5]	[0.6, 0.9]	[0.1, 0.5]
Propensity to be Selective	[0.1, 0.5]	[0.6, 0.9]	[0.1, 0.5]
Capacity to Store Messages	10	20	10
Initial Opinion	[0, .5] or [0.5, 1]	[0, 0.5] or [0.5, 1]	0.5

[j] This design of voter preference change is consistent with the idea of autoregressive influence[16] and conformity to the majority.[16,41] The concept of autoregressive influence refers to the influence of perceived external pressure, including peer pressure. Such social influence "depends on the distribution of opinion across all other individuals within the network who are also connected to the first individual";[4,p.20] in other words, when individuals perceive that messages from their social context turn to oppose their current preferences, they are likely to conform to the majority. In S-RAS, an agent whose current value in *Opinion* is higher than 0.5 will be recorded by the super monitor as 1 in *Voter Preference*, will be regarded by its network members as 1 in *Voter Preference*, and will also see itself as favoring 1 over 0.

The numbers in the brackets in Table 1 indicate the boundary for a normal distribution from which a random value is drawn. For example, for the propensity to discuss politics to vary from 0.6 to 0.9 means that S-RAS will draw a random number from a normal distribution bounded between 0.6 and 0.9 as an agent's *Propensity to Discuss Politics*. When simulation starts, every agent will have its unique *Propensity to Discuss Politics.*

4.5. *The context of political information*

Selective exposure is assumed in S-RAS. The contexts of political information for each agent in S-RAS include self-selected network members and self-selected news media. A communication network is composed of an agent and eight neighbor agents included on a 3*3 grid (a Moore neighborhood). Before simulation starts, each agent will form its own "contact list" based on the network members' *Political Expertise* and *Partisanship*. It is assumed that an agent will regard those (1) with greater political knowledge and (2) having congruent partisanship as the most favorable discussants.[45] Agents in S-RAS interact in a dyadic fashion. When an agent finds an available discussant, both agents will become unavailable to the other agents.[46] If political experts in the neighborhood are not available, the agent will turn to those with a lower level of political expertise but of the same party identification. Hence, on an agent's contact list the least favorable discussants are those with the lowest political expertise and those with different political party identification. Those agents placed at the bottom of the list are least frequently contacted.

A news media object in S-RAS refers to *any source of information other than dyadic interpersonal discussion*. By this definition, the news media objects include political elites who usually appear on TV, in newspapers, on the Internet, and in other kinds of news channels.[k] A news media object consistently holds a consistent voter preference. Unlike political discussant agents that may be unavailable sometimes, the news media objects can be accessed by any agent at any time. Hence, the two media objects in S-RAS—one favoring "1" and the other favoring "0"—can be seen as politically polarized media groups.[l]

That media objects are assumed to be polarized does not mean that they always broadcast 1 or always broadcast 0. It is further assumed that about two-thirds of chance agents obtain news messages that are consistent with their partisan

[k] Zaller suggests that the news media and political elites are one entity of information source, which means (1) that political elites usually exert their influence through TV, newspapers, radio, etc., and (2) that the news media, in effect, exert their influence by reporting news, talks, and activities of political elites. Following this perspective, S-RAS labels this media-elite entity as "the media" and focuses on the role of political experts.

[l] It is important to note that the news media objects, although I label them as "the news media," are not simply two TV stations or channels. They can be a TV channel, a newspaper, a radio program, a magazine, a news website on the Internet, or even a town hall meeting. Because the attention is put on what an individual agent actually receives, there is no need to create objects representing every specific type of public or private news source.

orientation. Therefore, agents, according to their propensity for selective perception, will obtain a variety of messages from the self-selected news source. An agent having voter preference "1" that is selective at a time step will, for example, store 1 in its memory at the time it accesses a news medium. If it is not selective at another time step, it will get whatever the news medium gives, either a 1 with a probability of 0.67 or a 0 with a probability of 0.33.

5. Alignment of Observable Processes

The above section describes how the RAS framework is operationalized and how S-RAS is created with additional assumptions derived from empirical findings regarding voter behavior. Although it is still early to say that S-RAS meets all requirements of internal validity, it does reveal a promise to align itself with a theory and to get embedded into empirical findings as much as it can.

Next, the attention is shifted to calibration, which is another important step to achieve external validity. "Even though we may be unable to model all the factors that limit the applicability of a theory, such factors must nevertheless be defined in a precise enough way to make them *amenable to empirical testing*".[35,p.157]

One way to align a theoretical model with observable processes is to customize the model with specific parameter values. For this project, data were collected before and after the 2006 Kaohsiung mayoral election (Dec 9, 2006) in Taiwan. The first wave was conducted between Dec 4 and Dec 7 ($N = 764$) and the second wave was conducted between Jan 20 and Jan 27 ($N = 650$). The data are weighted by population distribution across administrative districts — age, gender, and education level.

This election has a number of empirical features that are consistent with the design of S-RAS. First, as the descriptive statistics of the first-wave survey show, voters chose between two political parties — the Democratic Progressive Party (DPP) and the Kuomingtang (KMT). Candidates nominated by other marginalized political parties had almost no impact on the election results (only 1.3% of the total votes). Second, news media are generally polarized; voters can find channels supporting one side while others favor the other side. Third, most voters find like-minded people to discuss politics: 76.7 percent of respondents find that their discussants agree most of the time, while 78.8 percent find their favorite discussant supports the same candidate. Fourth, Kaohsiung voters go to a TV before discussing politics: 91.4 percent of respondents report that they obtain political news from TV and 52.9 percent of respondents will continue go to a TV for further information.

For the calibration of S-RAS, Table 2 lists four criteria to draw information from the first wave of the survey: the proportion of political experts, the proportion of political experts who favored the DPP, the proportion of ordinary voters who favored the DPP, and the proportion of ordinary voters who have no preferences about the political parties, at least said "it depends" or "I don't know" when they were asked

Table 2. System parameters for calibration.

Parameter Names	Description	Values[a]
probYES	the proportion of ordinal agents ($C1$) holding voter preference "1";	0.167 (0.5, 0.9)
probExperts	the proportion of expert agents ($C2$);	0.343 (0.1, 0.9)
probExpertYES	the proportion of expert agents ($C2$) holding voter preference "1";	0.521 (0.4, 0.6)
probNOIDEA	the proportion of "independent" agents ($C3$);	0.588 (0.1, 0.9)

Note: [a]parameter values are descriptive statistics of Kaohsiung voters collected before the 2006 mayoral election; in the parentheses are values for a robustness/sensitivity test (see Table 3 on page 21).

about the political party they tend to support. [m] These four criteria are easy to find in most surveys and can find their corresponding parameters in S-RAS. First, 16.7 percent of respondents who did not think of themselves as politically knowledgeable (probExpertYES) favored the DPP; the proportion of $C1$ agents favoring "1" (probYES) was set as 0.167. Second, 34.3% of the respondents subjectively thought that they knew more about politics than their political discussants; the proportion of $C2$ agents in S-RAS (probExperts) is hence set to 0.343. Third, 52.1% of those who thought of themselves as politically knowledgeable favorred the DPP; hence, in S-RAS I set the proportion of $C2$ agents holding voter preference "1" (probExpertYES) to 0.521. Fourth, 58.8% of respondents who felt less politically knowledgeable had no clear voter preference. Hence, the parameter value for probNOIDEA is set to be 0.588.

There are three concerns while performing this empirical calibration. Not specific to this project or ABM, they are indeed the major loopholes of empirical validation approach for disciplines that consider matching empirical data with laboratory results. First, there are approximately no assumptions held in the S-RAS model that will be supported by empirical data. For example, in S-RAS it is assumed, for reasons of simplicity, that all agents have eight discussants. Although the first-wave survey shows that most (84.3 percent) of Kaohsiung voter networks are about eight or less in terms of the number of discussants, some (15.7 percent) have more than eight discussants. Second, when constrained by interview time through the telephone, it is difficult to collect sufficient information for all parameters and to validate the assumptions used in S-RAS. Third and most importantly, it is questionable to believe that the information collected through a telephone survey reflects the true preferences of respondents.[25] In particular, given the atmosphere where Taiwanese voters mistrust pollsters and are not patient to answer questions regarding party

[m]Following an assumption that those who are politically knowledgeable should have developed some voter preferences, I therefore code as missing the politically knowledgeable who said "I have no preference," "It depends," and "I don't know".

orientation, it is questionable to assume that the information plugged into the parameters is correct.

6. Alignment of Outputs

When we are not sure about whether the models derived from a theory have the internal resources to specify their own domain of application, "models must be put in correspondence with the real world by means of a theoretical hypothesis stating what kind of relation holds between a given model (or set of models) and a given real-world system (or set of systems)"[35,p.156]

Not every theory or theoretical framework provides clear guidance for creating theoretical hypothese. The RAS framework in this project is an example. RAS gives *an explanation* instead of *ways to make predictions* regarding the formation of voter preference. Hence, in this project I try to employ a naive approach to align the outputs: present the results before and after the simulation and compare them with the descriptive statistics of the second-wave survey.

The alignment of outputs is presented in the order of (1) simulation results based on calibrated parameters, (2) robustness test or sensitivity test results derived from a series of simulations on the calibrated model, (3) a comparison across the two surveys and simulation.[n]

6.1. *Simulation results based on calibrated S-RAS*

This Kaohsiung model (i.e. the S-RAS model calibrated with descriptive statistics of the Kaohsiung election) is run for 150 time steps to simulate the short-term period of opinion dynamics. Figure 2 presents four snapshots of the 40×40 opinion grid at the 3rd, 20th, 50th, and 150th time steps. The gray scale for each of the 1,600 cells reflects

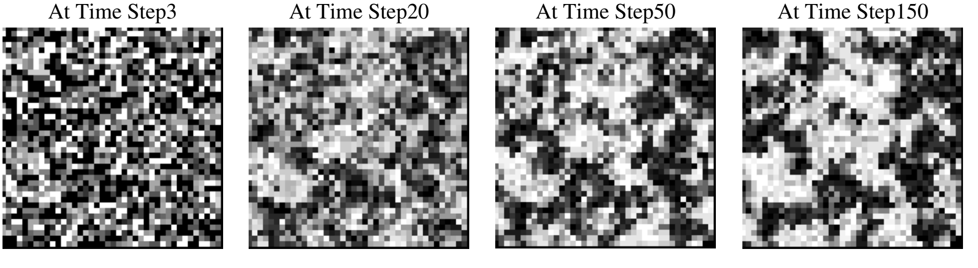

Fig. 2. Visualized distribution of opinion.
Note: The 40×40 grid of Moore neighborhoods is composed of 103 $C1$ agents, 543 $C2$ agents, and 954 $C3$ agents, who are randomly located. The opinions of agents are shown on a gray scale, where white denotes "0," and black denotes "1".

[n]If the purpose of the project is prediction, further calibration should be conducted based on the discrepancy found in (2) and (3), and the results of the simulation based on the re-calibrated model should be contrasted with (1). This process, however, has gone beyond the scope of this paper.

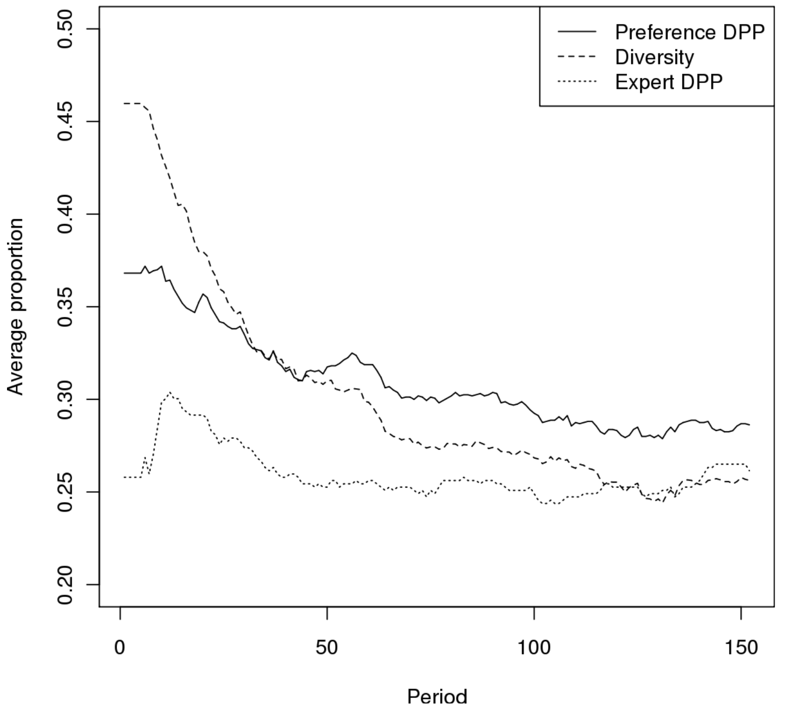

Fig. 3. The dynamics of preference.

the corresponding agent's opinion at the given time step. White denotes favoring the DPP ("1" or "YES" in the original design) and black denotes favoring the KMT (or "0" or "NO" in the original design). The formation of white and black clusters suggests a decrease in the number of independent voters and an increase in the number of opinion clusters.

Figure 3 shows the stabilization of voter preferences. The proportion of agents (including ordinary and expert agents) favoring the DPP decreases from 0.37 to 0.30. The level of diversity — measured by the average number of network members holding an opposite voter preference — also decreases, indicating the formation of homogeneous clusters. The proportion of expert agents favoring DPP increases a little bit in the beginning of the simulation but decreases and stabilizes at the level of 0.25.[o]

6.2. *The results of robustness tests based on calibrated S-RAS*

Before contrasting simulation results with empirical patterns, it is important to explore the extent to which the Kaohsiung model is sensitive to extreme parameter

[o]From a theoretical perspective, it is important to extend this pattern and then explore the conditions under which this pattern would change and remain. To an ABM modeler concerned with empirical validation, a more important question would be: what is the extent to which we take the above information and make inferences from the pattern?

Table 3. Summary statistics for the 100 runs of the simulation and robustness test.

Parameters	Default	probYES		probExperts		probExpertYES		probNOIDEA	
	-[a]	0.5	0.9	0.1	0.9	0.4	0.6	0.1	0.9
	Mean (S.D.)	Mean (S.D.)	Mean (S.D.)	Mean (S.D.)	Mean (S.D.)	Mean (S.D.)	Mean (S.D.)	Mean (S.D.)	Mean (S.D.)
probYES	0.299	0.514	0.567	0.361	0.549	0.398	0.515	0.258	0.518
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.05)
probExpertYES	0.258	0.516	0.570	0.336	0.551	0.380	0.524	0.226	0.518
	(0.04)	(0.06)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)
probNOIDEA	0.123	0.133	0.131	0.153	0.935	0.130	0.132	0.137	0.132
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Diversity	0.240	0.270	0.263	0.243	0.253	0.260	0.267	0.235	0.265
	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Extreme	0.262	0.254	0.255	0.201	0.372	0.257	0.254	0.264	0.253
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)

Note: [a]probYES = 0.167; probExperts = 0.343; probExpertYES = 0.521; probNOIDEA = 0.588.

values. Table 3 shows the influence of manipulated parameter values on five descriptive statistics, given that for each specific parameter value change the Kaohsiung model is run with 100 different random seeds (i.e. with 100 different random distributions in the first place). The five statistics to be observed include the proportion of agents favoring the DPP (probYES), the proportion of expert agents favoring the DPP (probExpertYES), the proportion of independent agents (probNOIDEA), the average number of discussants an agent perceives to hold opposite voter preferences (Diversity), and the average number of agents whose opinion is greater than 0.9 or lower than 0.1 (Extreme).

With all other parameter values being consistent with the default setting of the Kaohsiung model, the eight attempts to change parameter values generally do not result in extreme odd values. Plugging a higher value of probYES, 0.5, for example, (see the 2nd column) reasonably increases the consequent value of this parameter (0.51) but plugging an extreme value 0.9 does not significantly change this pattern. The proportion of agents favoring the DPP goes up mildly to 0.57.

The most important exception is that probNOIDEA is sensitive to the change in the value of probExperts. When probExperts is 0.343 as set in the default Kaohsiung model, probNOIDEA goes down to 0.123 when simulation stops. Lowering down the value of probExperts to 0.1 does not pull down the result value of probNOIDEA, but when probExperts is greater than 0.9, probNOIDEA goes up to 0.935. Among the five statistics, Diversity is the least sensitive to extreme parameter values (varying between 0.23 and 0.27).

6.3. *Comparison between survey results and simulation results*

Table 4 presents the comparison based on five parameters that can be found in both the simulation and survey questionnaire: probYES, probExperts, probExpertYES,

Table 4. Comparison between empirical and simulation patterns.

Parameters	First-Wave	Second-Wave	Simulation
probYES	0.167	0.196	0.299 (0.04)
probExperts	0.588	0.431	0.588[a]
probExpertYES	0.343	0.221	0.258 (0.04)
probNOIDEA	0.521	0.924	0.130 (0.01)
Diversity	0.537	0.568	0.240 (0.02)

Note: [a]S-RAS assumes that the number of political experts remains stable.

probNOIDEA, and Diversity. The first four are those used for calibration (recall Table 2). As discussed in the first half of the paper, there is not much reason to compare simulation results and empirical findings directly. It is, therefore, not a big surprise to see that the simulation results resemble the second-wave survey results. What we can inspect, however, is whether there is any trend found in the empirical data that is caught in simulation. Both the empirical survey data and simulation results suggest that during the six weeks (150 time steps) (1) the proportion of individuals supporting for the DPP increases, (2) the proportion of the politically aware decreases, and (3) the level of Diversity remains stable.[p]

Two discrepancies require attention. First, the proportion of (self-claimed) political experts decreases during the election, while in S-RAS this proportion is assumed to remain stable over time. This discrepancy can be a result of using the self-evaluation of political knowledge, instead of objective measures of political knowledge, in the surveys. Therefore, it will be important to explore whether the assumption "the level of political expertise remains stable" in S-RAS passes more empirical tests.

Second, the proportion of respondents saying they have no preferences in the second wave of the survey soars from 52 to 92 percent. S-RAS was not able to catch this trend and, instead, suggested that the proportion will decrease dramatically. A similar puzzle is found regarding the level of Diversity. Although parameter values remain stable, S-RAS suggests that the Diversity level will decrease instead of remaining at the 0.5 level. In other words, S-RAS exaggerates the tendency to form homogeneous opinion clusters.

7. Discussion

The spread of the ABM approach across disciplines is similar to the process of the adoption of statistical methods in both the natural and social sciences. It is therefore important to summarize the status quo for scholars paying attention to this approach.

[p]Note that in the survey questionnaire Diversity means that half or more than half of the discussants disagree when discussing politics, while in S-RAS Diversity refers to the average of the number of network members an agent perceives to hold opposite voter preferences. Although the measurement of Diversity differs, both empirical and simulation results agree that this parameter is a stable one. In empirical data, the portion falls between 0.53 and 0.57, suggesting that political disagreement does not diminish before or after the Election Day.

This paper is an initial attempt to narrow the gap between disciplines by summarizing how scholars across disciplines deal with the issue of empirical validation for computational simulations, before giving an example of the application of the principles to a study of the dynamics of voter preference. ABM scholars usually find that, while it is important "to communicate with colleagues who distrust the result of computer simulations," finding and presenting a successful empirically valid model and its results is a difficult and complex task.

Then, a dilemma comes to ABM modelers: on the one hand, it is important to make the design of the model transparent by listing all these assumptions, but this will make a paper more technical and more difficult to read; on the other hand, without explaining assumptions as much as possible, it will become difficult to persuade a skeptical academic community.

As Janssen and Ostrom suggest, if the data are substantial and of good quality, and if one can draw stylized facts from such empirical data, researchers can derive statistical distributions and other stylized facts from the empirical data; if the model is relatively uncomplicated, researchers can ask what are the simple rules that generate these stylized facts and then investigate the modeled conditions under which they can derive similar statistics to the observed styled facts.[19,q] Indeed, comparing statistical relationships based on empirical data and those found in a simulation is a more sophisticated approach than attempting to plug in some empirical values and make predictions. The discrepancy shown in this paper is an example of the latter.

What they have not pointed out, however, is when an agent-based model will be sophisticated enough for scholars to draw proper and testable hypotheses. To be specific, it is not clear to ABM modelers (1) which parameters should be used when a theory provides few hypotheses to test, and (2) how sure we are that an agent-based model is loyal to a theory. Without fully understanding a crystal ball or a black box created in the laboratory, it remains a challenging task for ABM scholars to conduct empirical validation.

The lesson I learned from this project is that a project aiming at a theoretical goal can be discouraged when taking into account the consideration of empirical validation, which requires much more detail to justify a model design. A prediction-oriented project that emphasizes details can also suffer from criticism for lacking internal validity. Looking to the future, ABM modelers, as well as scholars evaluating this method, need to continue to think about the issues surrounding external validation.

[q] One original goal of this paper was to conduct a series of comparisons of the following three hypotheses regarding political communication: (1) a voter who frequently discusses politics is likely to perceive less preference heterogeneity in his or her communication network; (2) a voter who frequently accesses news media is likely to perceive preference heterogeneity; (3) a voter who seldom discusses politics is likely to be ambivalent in its voter choice. The problems found during the time of writing this paper are not (1) these hypotheses are derived from the RAS framework, and that (2) the tests will become meaningless if all of the elements for internal validity have been fully examined. This goal, therefore, should be postponed to the next project after all necessary robustness tests are performed.

For scholars who like to make the S-RAS model presented in this paper more applicable for further tests using empirical data, there are two suggestions. First, assumptions held in S-RAS need to be further relaxed and calibrated, if they cannot yet be empirically tested. Calibration by changing parameter values, as demonstrated in this paper, is not enough, because this type of calibration should be based on a solid ground that the assumptions being held in the model are checked. For example, S-RAS assumes that all agents think about checking out news media before discussing politics. One will need to deal with issues like (1) what if this assumption is relaxed, (2) what if all agents reverse this process, or (3) to what extent will this assumption change affect the overall result? Second, there is a need for a step beyond a robustness test: to meet "generative standard" or "proximal similarity" — an assumption of generalization from concrete particular treatment operations.[22] As McQuarrie suggests, one important approach to promise proximal similarity is to use different combinations of parameters/variables chosen from the set of parameter/ variables drawn from the theory. Specifically, the goal of proximal similarity assumes that "if one member of a set of parameters $\{t_1 \ldots t_k.\}$ causes a certain outcome, a new sampling from that set will cause a similar outcome".[22,p.145] "We have to warrant to suppose that any other treatment operation drawn from the set $\{t_1 \ldots t_k.\}$ would also have a similar effect".[22,p.144] If more than one candidate can explain the phenomena of interest, as is often the case in computational modeling, "further work is required at the micro-level to determine which [model specification] is the most tenable explanation empirically".[47,p.43]

Appendix: Detailed Specification of the S-RAS Model

Table 5. System parameters.

Parameters	Description	Values
probExperts	the proportion of expert agents ($C2$);	0.343 (0.1, 0.9)
probNOIDEA	the proportion of "independent" agents ($C3$);	0.521 (0.1, 0.9)
probYES	the proportion of ordinal agents ($C1$) holding voter preference "1";	0.167 (0.5, 0.9)
probExpertYES	the proportion of expert agents ($C2$) holding voter preference "1";	0.588 (0.4, 0.6)
worldXSize, worldYSize	dimensions of the grid (the grid is wrapped as a torus);	40×40
radius	the "distance," i.e., the number of grid cells, from which an agent can reach other agents in 8 directions. If the radius equals 1, the size of a Moore neighbor will include 9 agents; if the radius equals 2, the size will be 25.	1
numMedia	the number of positions available from the news media;	2
duration	the total number of time steps for simulation;	150
memLength	the number of messages (0 or 1) $C1$ and $C3$ agents can process in each time step;	10

Table 5. (*Continued*)

Parameters	Description	Values
memLength2	the number of messages (0 or 1) $C2$ agents can process in each time step;	20
c1ExpertiseMin/Max	the lower/upper bounds of a random integer drawn from a normal distribution that is to be assigned to $C1$ and $C3$ agents as an indicator of expertise levels;	1, 5
c2ExpertiseMin/Max	the lower/upper bounds of a random integer drawn from a normal distribution that is to be assigned to $C1$ and $C3$ agents as an indicator of expertise levels;	6, 10
c1Conform	the threshold point of opinion at which $C1$ and $C3$ agents change their voter preferences;	0.5
c2Conform	the threshold point of opinion at which $C2$ agents change their voter preferences;	0.5
c1SPMin/Max	the lower/upper bounds of a random decimal number drawn from a normal distribution that is to be assigned to $C1$ and $C3$ agents as an indicator of the propensity to perform selective perception of news media messages;	0.1, 0.5
c2SPMin/Max	the lower/upper bounds of a random decimal number drawn from a normal distribution that is to be assigned to $C2$ agents as an indicator of the propensity to perform selective perception of news media messages;	0.6, 0.9
c1MPMin/Max	the lower/upper bounds of a random decimal number drawn from a normal distribution that is to be assigned to $C1$ and $C3$ agents as an indicator of the propensity to access a news media object;	0.1, 0.5
c2MPMin/Max	the lower/upper bounds of a random decimal number drawn from a normal distribution that is to be assigned to $C2$ agents as an indicator of the propensity to access a news media object;	0.6, 0.9
c1TPMin/Max	the lower/upper bounds of a random decimal number drawn from a normal distribution that is to be assigned to $C1$ and $C3$ agents as an indicator of the propensity to discuss politics;	0.1, 0.5
c2TPMin/Max	the lower/upper bounds of a random decimal number drawn from a normal distribution that is to be assigned to $C2$ agents as an indicator of the propensity to discuss politics;	0.6, 0.9
MEDIAONLY	if agents only access the news media for campaign information;	NO
GRAY	if a gray-scale is used for the color of the opinion grid (if NO, the grid will show a green-black color scheme);	YES
CHECKTALK	if agents follow their propensities to discuss politics;	YES
c1CHECKMEDIA, c2CHECKMEDIA	if $C1$, $C2$, and $C3$ agents follow their propensities to access the news media.	YES

Note: All features of $C3$ agents are the same as $C1$ agents except that $C3$ agents' voter preference is initiated as 0 or 1 at random and those of their opinions are initiated to be 0.5 instead of 1 or 0.

Acknowledgments

The first draft of this paper was presented in the American Political Science Association annual conference Aug 30−Sept 2, 2007 in Chicago, IL., USA. The author is grateful to the China Foundation for the Promotion of Education and

Culture for a travel grant and National Science Council for research grant (96-2414-H-110-001-). The source code of the program and survey data used in this paper are available from the author.

References

1. R. Axelrod, *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration* (Princeton University Press, Princeton, N.J., 1997).
2. L.-E. Cederman, Modeling the size of wars: From billiard balls to sandpiles, *American Political Science Review* **97**(1) (2003) 135–150.
3. J. N. Druckman, D. P. Green, J. H. Kuklinski and A. Lupia, The growth and development of experimental research in political science, *American Political Science Review* **100**(4) (2006) 627–635.
4. R. Huckfeldt, P. E. Johnson and J. D. Sprague, *Political Disagreement: The Survival of Diverse Opinions within Communication Networks* (New York: Cambridge University Press, 2004).
5. P. Johnson, Simulation modeling in political science, *American Behavioral Scientist* **42**(10) (1999) 1509–1530.
6. M. Laver, Policy and the dynamics of political competition, *American Political Science Review* **99**(2) (2005) 263–281.
7. F. C. Liu, Constrained opinion leader influence in an electoral campaign season: Revisiting the two-step flow theory with multi-agent simulation, *Advances in Complex Systems* **10**(2) (2007) 233–250.
8. I. S. Lustick, D. Miodownik and R. J. Eidelson, Secessionism in multicultural states: Does sharing power prevent or encourage it? *American Political Science Review* **98**(2) (2004) 209–229.
9. J. M. Epstein and R. Axtell, *Growing Artificial Societies: Social Science from the Bottom Up*, Complex Adaptive Systems (Cambridge, MA: The MIT Press, Washington, D.C., 1996).
10. G. N. Gilbert and K. G. Troitzsch, *Simulation for the Social Scientist* (New York: Open University Press, Buckingham; Philadelphia, Pa., 2005), 2nd edition.
11. M. M. Waldrop, *Complexity: The Emerging Science at the Edge of Order and Chaos* (Simon & Schuster, New York, 1992).
12. A. Schram, Artificiality: The tension between internal and external validity in economic experiments, *Journal of Economic Methodology* **12**(2) (2005) 225–237.
13. R. Sun, ed., *Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation* (Cambridge University Press, New York, 2006).
14. R. M. Axelrod, *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*, Princeton studies in complexity. (Princeton University Press, Princeton, N.J., 1997).
15. D. Rousseau and A. M. Van der Veen, The emergence of a shared identity: An agent-based computer simulation of idea diffusion, *Journal of Conflict Resolution* **49**(5) (2005) 686–712.
16. R. Huckfeldt and J. M. Mendez, Managing political heterogeneity: Perceptions of disagreement within communication networks, (Paper presented in the Midwest Political Science Association 62nd Annual National Conference, April 15–8, 2004, Chicago, IL, 2004).
17. D. G. Brown and D. T. Robinson, Effects of heterogeneity in residential preferences on an agent-based model of urban sprawl, *Ecology and Society* **11**(1) (2006) 46.

18. R. Garcia, P. Rummel and J. Hauser, Validating agent-based marketing models through conjoint analysis, *Journal of Business Research* **60**(8) (2007) 848–857.
19. M. A. Janssen and E. Ostrom, Empirically based, agent-based models, *Ecology and Society* **11**(2) (2006) 37.
20. G. Laurent, Improving the external validity of marketing models: A plea for more qualitative input, *International Journal of Research in Marketing* **17**(2–3) (2000) 177–182.
21. J. W. Lucas, Theory-testing, generalization, and the problem of external validity, *Sociological Theory* **21**(3) (2003) 236–253.
22. E. F. McQuarrie, Integration of construct and external validity by means of proximal similarity: Implications for laboratory experiments in marketing, *Journal of Business Research* **57**(2) (2004) 142–153.
23. K. Mehta and S. Bhattacharyya, Design, development and validation of an agent-based model of electronic auctions, *Information Technology and Management* **7**(3) (2006) 191–212.
24. D. F. Midgley, R. E. Marks and D. D. Kunchamwar, The building and assurance of agent-based models: An example and challenge to the field, *Journal of Business Research* **60**(8) (2007) 884–893.
25. J. Zaller, *The Nature and Origins of Mass Opinion* (Cambridge University Press, New York, 1992).
26. B. McKelvey, Model-centered organization science epistemology, in *The Blackwell Companion to Organizations*, ed. J. A. Baum (Blackwell Publishers, Malden, Mass, 2002), pp. 752–780.
27. J. Becker, B. Niehaves and K. Klose, A framework for epistemological perspectives on simulation, *Journal of Artificial Societies and Social Simulation* **8**(4) (2005) http://jasss.soc.surrey.ac.uk/8/4/1.html.
28. A. Schmid, What is the truth of simulation? *Journal of Artificial Societies and Social Simulation* **8**(4) (2005) http://jasss.soc.surrey.ac.uk/8/4/5.html.
29. R. M. Entman and S. Herbst, Reframing public opinion as we have known it, in *Mediated Politics: Communication in the Future of Democracy*, eds. W. L. Bennett and R. M. Entman (Cambridge University Press, New York, 2001), pp. 203–225.
30. D. A. Graber and J. M. Smith, Political communication faces the 21st century, *Journal of Communication* **55**(3) (2005) 479–507.
31. J. S. Lansing, Complex adaptive systems, *Annual Reviews of Anthropology* **32** (2003) 183–204.
32. J. H. Kagel and A. E. Roth, eds., *The Handbook of Experimental Economics* (Princeton University Press, New Jersey, 1995).
33. G. Fagiolo, P. Windrum and A. Moneta, Empirical validation of agent based models: A critical survey, *Working Paper* (2006) http://www.lem.sssup.it/WPLem/files/2006–14.pdf.
34. A. Nowak, J. Szamrej and B. Latane, From private attitude to public opinion: A dynamic theory of social impact, *Psychological Review* **97**(3) (1990) 362–376.
35. F. Guala, *The Methodology of Experimental Economics* (Cambridge University Press, New York, 2005).
36. J. N. Druckman and A. Lupia, Preference formation, *Annual Review of Political Science* **3** (2000) 1–24.
37. M. B. Oliver, Individual differences in media effects, in *Media Effects: Advances in Theory and Research*, eds. J. Bryant and D. Zillmann, 2nd edn. (L. Elbaum Associates, Mahwah, NJ, 2002), pp. 507–524.
38. P. Converse, *The Nature of Belief Systems in Mass Publics* (Free Press, New York, 1964), pp. 206–261.

39. J. Jerit, J. Barabas and T. Bolsen, Citizens, knowledge, and the information environment, *American Journal of Political Science* **50**(2) (2006) 266–282.
40. D. C. Mutz and P. S. Martin, Facilitating communication across lines of political difference: The role of mass media, *American Political Science Review* **95**(1) (2001) 97–114.
41. W. R. Neuman, M. R. Just and A. N. Crigler, *Common Knowledge: News and the Construction of Political Meaning* (University of Chicago Press, Chicago, 1992).
42. L. M. Bartels, Unenlightened self-interest, *The American Prospect* **15**(6) (2004) A17–A19.
43. P. Converse, Popular representation and the distribution of information, in *Information and Democratic Processes* (University of Illinois Press, Urbana, IL, 1990), pp. 369–90.
44. M. P. Fiorina, S. J. Abrams and J. Pope, *Culture War? The Myth of a Polarized America* (Pearson Longman, New York, 2005).
45. R. Huckfeldt, The social communication of political expertise, *American Journal of Political Science* **45**(2) (2001) 425–438.
46. R. Huckfeldt and J. D. Sprague, *Citizens, Politics, and Social Communication: Information and Influence in an Election Campaign* (Cambridge University Press, England, 1995).
47. J. M. Epstein, Agent-based computational models and generative social science, *Complexity* **4**(5) (1999) 41–60.