Abstract—Information system development is a result of a complex mix of forces that are controlled by technological issues as well as social, organizational and economic issues. Starting from the need to compare the long term effect of the centralized and decentralized systems and more particularly, the utility of the swarm executable choreographies, the present paper comes up with a method of analyzing the effects which the information system may have over society. Our analysis is not meant to predict some precise values over some indicators but rather to predict the value of some comparative measures among indicators. The events taken into account will be the Black Swan events and sequences of complex events whose effects are difficult to be estimated without software methods.

Keywords - ISD impact modeling; privacy risks; Monte Carlo method; social effects estimation; centralised and decentralised systems

I. INTRODUCTION

This article is intended to be related to the methods of a relatively new science Cliodynamics [1, 2] and suggests a multidisciplinary solution in order to analyze the effects in time of the decisions related to the development, implementation and usage of the software systems.

Thus, the method suggested here will be entitled INSIGHT. With this method, it will be possible to give arguments related to the decision to implement and use software or a certain architecture. Our previous research in the field of the executable choreographies encounter a problem: how could it be proven that the software architectures and the methods meant to integrate the systems and the web services using executable choreographies [3] may be superior to some methods based on centralized methods such as the orchestration of web services [4]. The empirical observations resulted from the practical use of executable choreographies have intuitively indicated that their use on a large scale may have multiple uses. These are meant to ease programmers’ life on the one hand and to identify some social effects which are not visible at first sight. Although the future effects of a certain technology may hardly be numerically estimated, this article has as main purpose to objectivize the effects that technology will have over society. The recommended method intends to start from experts’ experience and intuition in creating “educated guesses” about future events. We shall focus mainly on very important events, for example Black Swan events [5]. Being so rare, these events are usually underestimated, but their consequences are very powerful. The suggested method is based on the experts’ evaluation regarding possible events, regarding probabilities and the magnitude of these events. Making a numerical estimation, it is possible to make comparisons regarding different solutions and scenarios. The method envisaged the result of some numerical indicators or the use of a certain solution. Practically, it may be used in order to analyze some long term decisions or actions generally speaking. Such an indicator may be the estimation of some financial gaining, or more seriously, the number of some victims or the number of the people involved in an accident. Our interpretation considers relevant the differences in value between two or more solutions. It is important to underline that the exact number is not a target, but rather an estimated value is what we are concerned with. It is believed that using an indicator such as Happiness Indicator [6] and the estimation of such index may lead to the use of reliable emotional values which could guide the decision takers. For instance, based on a prediction of a happiness indicator or on the numbers of victims, it is possible to get reliable predictions of the effects caused by one solution or another. As it is going to be demonstrated later on, if it is applied to a series of events which may influence the population of a country or of the entire world. Some decisions may have significant consequences by comparison.

II. INSIGHT METHODE

The INSIGHT method of estimating the effects of future events consists in the gathering some relevant events in the estimated field and some numerical figures obtained from the specialists in the field. The beliefs, events and other rules coming from the specialist turn into scripts, entitled INSIGHT which is executed until it is possible to determine the effects of the estimation. The Method INSIGHT suggests taking the following steps:
• Start informal discussions among the experts who will make the estimation. Improve the intuition of the participants on the subject meant to be analyzed. It is essential that all experts involved in deciding should understand similar historical situations and to be aware of the solutions of possible chains of events.

• The second step comprises the formal definition of the comparison objects. As we are not interested in accurate predictions, but rather in estimating the scale of the effects, our approach necessarily implies a comparison of solutions or scenarios. The comparison may concern the effects of two solutions or the effects of different scenarios concerning the same solution. By solution we define the specific approach of a problem, while the scenario describes sets of possible events for each solution. For example, a centralized or decentralized system may be considered a solution, while the scenario may be a perspective involving a succession of pessimistic or optimistic events. Thus, it is mandatory to define at least two solutions or two scenarios. This step concludes by compiling a list of relevant scenarios concerning the compared solutions. Subsequently, we are to detail an example wherein we compare two scenarios, as well as one comparing two solutions.

• The third step of the method consists in determining an event list directly or indirectly dependent on the assessed solutions and scenarios. Furthermore, the experts are to issue laws able to modify the beliefs in the probability of certain events as the consequences of another events. The dependency is not to be interpreted as strict causality, but rather as the effects of a catastrophic or auspicious independent event may occur in the future and may affect events in possible universes. For instance, we may conduct analysis on the effects of a dictatorship in a specific country enabled to peruse the users’ data of a widely used social network. Although the dictatorship is not necessarily directly influenced by the existence of the social network, the data available in the aforementioned network may severely affect the citizen’s life, as they are prone for economic, social and political discrimination. We will exemplify the assessment of such a use-case in the 4 section.

• The fourth step consists in estimating the participating experts’ beliefs regarding the analysis of each assessed event. The belief is a value between null and one and represents an intuitive option of an expert assigned to the probability of the prediction occurrence in one of the following years. One must observe the difference between the probability of occurrence of an event in a random year and the occurrence of an event in one of the years covered by the simulation. The INSIGHT method assumes the experts are estimating the effect is to happen in one of the following years. The effects’ scale of magnitude considered by the simulation may be greatly influenced by the year the event is supposed to occur. As we are about to demonstrate, the simulation tries to generate based on laws issued by the experts a number of variants concerning the studied scenario as high as possible, in order to achieve a mean effect as viable as possible. In many cases of our simulations, a Black Swan effect occurring in an immediate year generates a much more potent effect than one occurring in the distant years of the simulation. As such, the year is paramount for the accuracy of the estimation, and even more for the comparative amplitude of the simulated scenarios.

• The fifth step consists in estimating the laws of acquiring the events’ effects. For instance, the experts may issue percentage of users directly affected by an event or other complex effect determination rules.

• The sixth step consists in coding the probabilities and rules acquired by the expert. We are suggesting a specific language we will refer to as INSIGHT, which is already implemented as open source [7]. The simulation is saved in a text file, enabling further reviewing, adjusting and improving of the simulation. Our approach implies using an intern DSL (Domain Specific Language) [8] for the JavaScript comprising a series of specific sections covering:
  
  o The description of the simulation variables. Every simulation owns a series of variables customizable during the simulation. For instance, the simulations may have variables describing the number of affected population, the number of casualties, happiness indexes of temporary variables used as probability estimation rules of event effects estimation rules.

  o General environment settings for the simulation (independent of the scenarios). For instance, we may set the number of years or the minimum or the maximum stochastic executions for a scenario (maxSimulations, minSimulations), as well as the distribution type.

  o In order to compare two simulations or two solutions, it is necessary to consider the same number of years for all the simulations and scenarios. In our implementation we used four categories of event distribution:
    - The Gauss distribution also known as the normal distribution.
    - The uniform distribution enabling random occurrence of the events in the considered period.
A special “median” distribution generating events years only in the middle of the simulated intervals, enabling faster computing and being closer to the results acquired by an expert using only a general prediction software. Currently, if the experts are unable to enlist a programmer to make use of specific instruments, they may still be able to employ for risk or effect assessment a general instrument such as a simple spreadsheet. The implementation is making use of the “median” distribution in order to enable experts to compare INSIGHT script implementation with their usual prediction methods.

- A special “certainly” distribution enabling simulations to be completely independent of experts’ beliefs. The “certainty” distribution is used to validate the experts’ intuitions, for comparisons and in order to simplify the scenarios where it is impossible to issues realistic beliefs.

- The description of scenarios and of the rules by which the scenarios may affect the events. Each scenario is to hold a function to be executed in the beginning of the simulation, enabling the modification of probabilities.

- The description of human beliefs in executing an event. In order to respect the intuition, the beliefs will be coded by simple numbers from 0 to 1, implicitly associated to a starting year (a year representing the simulation step) and an ending year representing the conclusion of the simulation. The events’ distribution will depend on the start and the end year. For instance, we will consider that the \( X \) event holds 60% (or 0.6) chances of occurrence during the simulation. In this fashion, the experts may easily add values for estimating a certain event.

- The description of the effects of an event on the simulation variables. This description will be carried out by executable functions, namely the “Effect” functions. They are to be mandatorily found in the description of each event.

- The description in executable code of the effects the passing of a year incurs on the current simulation variables. This description will be coded by two \( \text{beforeEachYear} \) and \( \text{eachYear} \) functions that are to be executed on each step of the simulation.

The final step is to execute the INSIGHT script, adjust the thresholds and to analyze the meaning of the results as it is exemplified in the following chapters. The simulation variables are numeric values in general, such as the affected population for instance. The INSIGHT system enables the post processing of the simulation values (values spread over years) by extracting the sum, average of maximum values.

### III. The Proposed Estimation Computing Method

The exact methods to acquire the numeric values of experts’ beliefs on the probability of future events is not the concern of this study. There is possible to employ historical data, at times, as there are also cases in which it is impossible to use valid historical data, especially when taking into consideration the Black Swan [2] events, or events implying new technologies without historical precedent. In cases such as these, there will be only the experts’ predictions to compute.

The concrete calculus for estimating a simulation’s effects is currently based on a stochastic method from the Monte Carlo [9] simulations family.

We are subsequently describing the algorithm implemented through the INSIGHT method:

- A lesser than one positive numeric value is assigned to a \( T \) variable called threshold. Usually, \( T \) receives a \( 1/\maxSimulations \) value.

- A high number of simulation cases is subsequently set. A simulation case represents an \( n \)-uple generated by the association of a simulation event with the year it is supposed to occur. We will subsequently refer to these events strings as simulation cases or shortly – cases. An execution year is generated for each case and for each possible event \( E \) in the current simulation, considering the events’ distribution as it is set up in the INSIGHT script. Depending on the belief \( B \) issued by the experts, the step is probabilistically assigned to become \(-1\) (meaning impossible) with the \( 1-B \) occurrence probability. The event in the \(-1\) year may not occur.

- The simulations are run for each possible case computing predictions by effectively running the simulation and by executing the years and events associated functions (the \( \text{beforeEachYear} \), \( \text{eachYear} \), and \( \text{Effect} \) are executed for each occurring event) in yearly succession. If the executed functions are changing the occurrence belief of an event, then the execution year of the aforementioned event is dynamically reassigned.

- The predictions mean value is computed for each possible case and an associated effect is acquired for each year of the simulation and for each simulation variable.
Depending of the variable meaning, the effect is stacked by executing a mean value or a sum of all the current values in a case.

The scenarios cases generation is concluded when the current mean value of the \( i \) step, represented by the \( M(i, v) \) notation, is described by the \(|M(i, v) - M(i-1, v)| \leq T\) formula for all the values of the \( v \) variables in the simulation.

### IV. INSIGHT METHOD APPLIED TO ASSESS THE EFFECTS OF A SOCIAL NETWORK ON USERS’ LIFE

The INSIGHT method will be employed as an example, in order to analyze the possible social effects incurring when using a social network, without experts’ generated numbers being available. It is important to notice that a social network acquiring data from an entire country or a huge portion of the human populace are examples of highly centralized systems. For instance, we will consider simulating the potential effects of a social network with a billion users. The data shared in the social network may be used to create a political profile of a particular citizen and be used as such against the concerned individual.

We are considering the following possible events:

- **Dictatorship**: an autocratic regime is established in the considered country or its policies are tightened.
- **Retaliation**: the autocratic regime is employing the data acquired through the social network to retaliate against the opposing categories of citizens or those prone to dissidence. May it be that 0.1% of that nation’s citizens are punished in the following years.
- **Restauration**: In order to increase the simulation’s complexity, we will insert in the simulation the event through which the autocratic regime is ousted after a number of years or it changing its policy and it does not employ the social network data anymore.

The experts and programmers may generate the following INSIGHT script:

```insight
var sim = {
    years:20,
    maxSimulations:10000,
    minSimulations:100,
    distribution: "normal",
    initial values of the simulation are assigned.
    A minimum of 100 and a maximum
    10000 scenarios are to be generated.
    The distribution is normal.
    The simulation will cover 20 years.
    The simulation variables are initialized.

    Variables:{
        annualCausalities:0,
        Casualties:0,
        People:1000000000
    },

    Scenarios:{
        Pessimistic:function(){
            this.UsagePercent = 0.3;
            this.AtRisk = 0;
            this.RiskPercent = 0.3;
            this.setBelief("Dictatorship",0.2);
        },
        Optimistic:function(){
            this.UsagePercent = 0.3;
            this.AtRisk = 0;
            this.RiskPercent = 0.1;
            this.setBelief("Dictatorship",0.1);
        },
    }
}
```

An Optimistic and a Pessimistic scenario are described, as well as how each affects the initial variables. The change of the RiskPercent variables and the belief in the “Dictatorship” are to be observed.
beforeEachYear: function()
{
    if(this.UsagePercent < 0.8)
    {
        this.UsagePercent += 0.1;
    }
    this.AtRisk = Math.floor(this.People * this.UsagePercent * this.RiskPercent);
    this.annualCausalities = 0;
},
eachYear: function()
{
    //console.log("Running action for year", this.currentYear(), this.inDictatorship);
    if(this.inDictatorship)
    {
        this.annualCausalities += Math.floor(0.01 * this.AtRisk);
    }
    this.Casualties += this.annualCausalities;
},

Events: {
    Dictatorship: {
        Description: "Major change in the government.",
        Belief: 0.4,
        distribution: "normal",
        Effect: function()
        {
            this.inDictatorship = true;
            this.setBelief("Retaliation", 0.8);
            this.setBelief("Restoration", 0.9);
        }
    },
    Retaliation: {
        Description: "Use of private data to punish disobeying citizens",
        Belief: 0.1, //it could happen even without Dictatorship
        Effect: function(currentScenario, history)
        {
            this.annualCausalities += Math.floor(0.01 * this.AtRisk);
        }
    },
    Restoration: {
        Description: "Return to democracy. End of",
        Belief: 0,
        Effect: function(currentScenario)
        {
            this.inDictatorship = false;
        }
    }
}

var res = h.run(sim);
console.log("Estimated total number of causalities in Pessimistic case:", res.total("Pessimistic", "annualCausalities"));
h.print(res, "Pessimistic", "Casualties");
h.print(res, "Pessimistic", "annualCausalities");
console.log("Estimated total number of causalities in Optimistic case:", res.total("Optimistic", "annualCausalities"));
h.print(res, "Optimistic", "Casualties");
h.print(res, "Optimistic", "annualCausalities");
The chart of the Figure 1 was generated as result of the simulation run. The effects difference between the pessimistic (red) scenario and the optimistic (blue) one is apparent.

The above presented numbers are not correspondent to initial values assigned by experts, but rather generated based on plausible events and values employed in demonstrative purposes. It is easily noticeable that the numbers generated in a scenario are significantly higher than those generated by the other. We believe that the importance of INSIGHT resides not in the numerical values acquired, but rather in the effect scale as well as in the absolute difference between the solutions and scenarios. In the demonstrative example provided in Figure 1, it is apparent that the pessimistic scenario outputs significantly higher values that the optimistic scenario, which intuitively validates the computing methodology.

V. INSIGHT METHOD APPLIED TO ASSESS TWO SOFTWARE ARCHITECTURES

The second example of this study uses for input our evaluations as security and distribution systems experts. The INSIGHT method will be employed for modelling the potential effects of implementing an informatics system based on a centralized architecture or a decentralized one (such as executable choreographies). The purpose of these estimations is purely demonstrative for the suggested method. As such, we will not consider a real country, but rather an ideal one, and we will conduct a comparative risk assessment of massive data loss impacting the national security of an imaginary country with 100 million citizens. We are basically referring to the SOCIAL risk category approached in [11]. SOCIAL risk area includes risks associated with unauthorized access to sensible data concerning national security or the security of a large community. The medical data of every citizen of a town, region or country is obviously an example of such a special risk category.

The centralized systems are prone to multiple attack directions (servers, corruptible individuals). As such, the probability for an information attack to succeed is much higher than in the case of a totally decentralized system. On the other hand, a decentralized system is more vulnerable to attacks viewing discrete parts of the system (by corrupting individuals or corrupting infrastructure). Once the increment in locations and corruptible or incompetent individuals, we considered the subsystems’ vulnerability to locally increase at its turn. Let us consider the decentralized system comprises 100 autonomous subsystems, each component holding the data of 1/100 of the nation's’ denizens. The centralized system comprises a single system holding all the data.

<table>
<thead>
<tr>
<th>INSIGHT Script</th>
<th>Section description</th>
</tr>
</thead>
<tbody>
<tr>
<td>var sim = {</td>
<td>General configuration of both scenarios.</td>
</tr>
<tr>
<td>years:10,</td>
<td></td>
</tr>
<tr>
<td>maxSimulations:10000,</td>
<td></td>
</tr>
<tr>
<td>minSimulations:1000,</td>
<td></td>
</tr>
<tr>
<td>distribution:&quot;uniform&quot;,</td>
<td></td>
</tr>
<tr>
<td>Variables:{</td>
<td>General configuration of the simulation variables for both scenarios. It is given a</td>
</tr>
<tr>
<td>People:100000000,</td>
<td>population of 100 million people and a 1% affected citizens.</td>
</tr>
<tr>
<td>harmedPercent:0.1,</td>
<td>The Harmed variable is assigned as termination condition for the simulation.</td>
</tr>
<tr>
<td>Harmed:0</td>
<td></td>
</tr>
</tbody>
</table>
Scenarios:

Centralised: function() {
    this.originalBelief = 0.01;
    this.subsistems = 1;
    this.setBelief("Breach", this.originalBelief);
},

Decentralised: function() {
    this.originalBelief = 0.01;
    this.subsistems = 100;
    this.setBelief("Breach", this.originalBelief);
},

beforeEachYear: function(history, currentYear) {
    this.Harmed = 0;
},

eachYear: function(history, currentYear) {
    //empty function – nothing to do in this simulation
},

Events: {
    Breach: {
        Description: "Major breach affecting all the data."
    }
}

Effect: function(currentScenario, history) {
    var howMany = Math.floor(Math.random() * (this.subsistems - 1)) + 1;
    this.Harmed += howMany * this.harmedPercent * this.People / this.subsistems;
    this.setBelief("Breach", this.originalBelief);
}

var res = h.run(sim);

The description of the two scenarios. A single category of event – “Breach” – is simulated, meaning a security error. The single difference is the different number of subsystems. The centralized scenario has one system, while the decentralized one has 100 subsystems.

The beforeEachYear function is only setting the initial victims value in the current year in the case of null scenario simulation.

The Effect function computes differently the effects (the number of affected persons) of the Breach event, depending on the number of subsystems.

The setBelief function called in eachYear enables the event even if it has happened before.

By running the two scenarios, the value of the Harmed variable in the Decentralized scenario is half the value in the Centralized one. This value is intuitive, of course, and it is acquired by processing the simulation input and it is consistent with the Law of large numbers [12].

We believe such simple examples are useful in validating the implementation, as it proves to be consistent with our intuition concerning the result. It also supports our faith in the method when employed in cases where making use of manual computing or spreadsheets are no longer possible, and the execution of a large number of simulations is required, specific to the Monte Carlo methods.

VI. THE ALGORITHM ACCURACY AND THE PREDICTION ACCURACY

Definition 1: We consider a prediction method as convergent if any scenario execution always concludes for a given T threshold (T is defined in the 3rd section).

Definition 2: We consider a prediction method as convergent to a V value represented by an n-uple \((V_1, V_2, \ldots, V_n)\) if every value acquired as result of every simulation run that outputs values for \((v_1, v_2, \ldots, v_n)\) variables is consistent with the condition \(|v_i - V_i| < T\), for i from 1 to n.

Theorem 1: The INSIGHT prediction method is convergent, if all the values of a \(v\) variable are modeling real social data and if all the values of the same variable \(v\) are keeping their sign.
Intuitive argument: The condition requiring the variables to keep their sign means that semantically, the variable represents a cumulative concept (casualties’ number, affected people, etc.). As such, the variable may acquire for a given simulation only positive or negative values. This situation is the most probable, as the simulations running the INSIGHT method aim at predicting cumulative effects.

Demonstration: We observe that the simulation’s functions are modeling real social data. As such, they are returning a maximum value and a minimum value for each variable. Generally speaking, the implemented rules will reflect relevant social phenomena, and we may conclude that |v| is finite and limited to the number of affected people. Let V represent the maximum value of |v| (absolute value) for any v and let M(i, v) be defined as in the 3rd section.

Thus, in order to demonstrate the propriety of convergence is sufficient to show that for any v variable there is an i, natural number, consistent with the following relation: |M(i, v) - M(i-1, v)| <= T.

Because |M(i, v) - M(i-1, v)| <= 2·V/i, it is obvious that an i exists for which 2·V/i <= T, given any V and T.

Theorem 2: The INSIGHT prediction method is convergent for any set T, and given a sufficiently high number of scenarios, it converges to the mean value of all possible scenarios. When defining a sufficiently high number of scenarios, we are consistent with the interpretation according to the law of large numbers.

Demonstration: For any i, there is an n and a p such as i = n + p, where n is the number of negative values of v variable, and p is the number of positive values.

Therefore, M(i, v) = (Sum(n) + Sum(p))/i, where Sum(k) is the sum of values for the k scenarios. |M(i, v) - M(i-1, v)| may be also described as |Sum(n) + Sum(p) - Sum(n-1) - Sum(p)|/i or |Sum(n) + Sum(p) - Sum(n) + Sum(p-1)|/i.

Both |Sum(n) - Sum(n-1)|/i, and |Sum(p) - Sum(p-1)|/i are limited by 2·V/i, similar to the 1st theorem.

Intuitively, the law of large numbers applied to Monte Carlo models returns a fixed ratio between n and p (not exactly fixed, but rather inserted into a certain interval), if we generate sufficient i scenarios for a given simulation (more specifically, a scenario in an S script).

Given this limited ratio between n and p, we may conclude that the M(i, v) mean values will also converge to a value close to the theoretical limit of all possible values means. We therefore have demonstrated that the INSIGHT method is accurately approximating the intuitive value of the means of all possible scenarios predictions.

Notes: The correct calculus of the difference between the ideal value of the mean and the value returned by convergence towards a given T threshold is difficult. We may, however, consider this value as being limited by the maximum mean value multiplied by T. In order to compute the variables mean maximum value, abstract interpretation [13] methods could be employed on the script code.

T is less than one and may be assigned a very small value. As such, from a practical perspective, the difference between the theoretical mean value and the effectively returned value may be decreased indefinitely.

As such, practically speaking, the presented method may be approximated with sufficient accuracy by effectively running the code with different T values. Successive simulations runs may also indicate if it is both viable and necessary to undertake additional executions with T values significantly decreased.

VII. CONCLUSIONS

This article is not aiming at giving real predictions on using social networking or employing centralized / decentralized architectures. We are rather suggesting an approach as scientific as possible in order to usefully predict the social impact of certain scenarios, and thus help in decision making. In developing informatics systems, as well as in the manner we use them, there are many important decision to take that are depending on future event strings on which we have precious little information. The method described in this study is offering a computational approach to reach better intuitive understanding of the effects these future event may trigger.

Although the Monte Carlo simulation method can require lots of programming, it could extend the relatively narrow limits of classical inference method regarding complex social phenomena. The Monte Carlo variant proposed in this article, the INSIGHT method and the INSIGHT language can decrease the programming efforts because the general simulation algorithms are reusable. Only the specific numbers, events and rules should be entered in a Domain Specific Language, substantially reducing the programming costs.

The impact and the importance of the practical effects of such events emphasize the significance of an estimation even though such estimations may be made only through the educated guess of the specialists belonging to different fields. The experts may use historical data but the purpose of this method is to shape some events for which there are not enough historical data because the studied phenomena are extremely new (take as an example a brand new software system).
The method is not aimed at acquiring maximum or minimum values for the effects, but focus on returning mean effect values based on beliefs and convictions provided by experts. We believe that should we achieve moderate numerical values for the effects, the decisions may be understood and accepted by people as being based on rational foundations of solutions and scenarios assessment. The INSIGHT method implies that for numerous practical situations, the mean values based predictions between two or more solutions, are returning results sufficiently different to justify taking a decision. Subsequent to the academic validation of the method, we intend to pursue this research trend by establishing multidisciplinary teams also involving humanities specialists. We will attempt to predict in this manner the social effects of employing certain software systems.

REFERENCES