Argumentation and belief updating in social networks: a Bayesian approach

Masterton, George; Olsson, Erik J

Published in:
Trends in belief revision and argumentation dynamics

2013

Argumentation and Belief Updating in Social Networks: A Bayesian Model

George Masterton Erik J. Olsson

September 10, 2013

Abstract

Human rationality involves arguing for one's beliefs as well as revising those beliefs in the face of good reasons to do so. An important problem in computer science and philosophical logic is to find models that combine argumentation and belief updating in one formal framework. In this article, we present the Laputa model which attempts to do precisely this. Laputa differs from some other frameworks with similar goals (a) in being Bayesian, (b) in taking *pro et contra* argumentation as the paradigm case, and (c) in taking the persuasive effect of an argument, rather than its strict acceptability, to be the more fundamental notion. This article provides a general introduction to the model for a computer science audience as a way of integrating argumentation and belief updating. It features detailed derivations of how inquirers in a social network update their credences and trust assessments in response to inquiry and argumentation.

1 Introduction

Human inquiry has many facets of which argumentation and belief revision stand out as particularly salient. Unsurprisingly, both have been studied extensively in various areas of scientific research. Computer science and philosophical logic are no exception to the rule. Both argumentation and belief revision have received extensive treatment by computer scientists and logicians alike. Yet, in standard treatments these aspects of rational inquiry are investigated separately rather than in conjunction. For example, most models of belief revision leave argumentation out of the picture; which is true, for instance, of the influential AGM model (Alchourrôn et al., 1985). This is so for a good methodological reason: both areas are quite complex in themselves and it makes good methodological sense to adopt a simplified picture for the time being. Still, the ultimate goal must surely be a model which incorporates both these phenomena. How this could be accomplished is the subject of this article.

There are many ways in which one could incorporate a role for arguments in belief revision. Belief revisionists realised early on that not all beliefs have the same status. Rather, some beliefs are held because of other beliefs: the latter provide the reasons for the former. A straightforward way to capture this phenomenon in a belief revision framework is to introduce a belief base into the model (e.g. Furhmann (1991), Hansson (1991)). The belief base of an agent contains all the beliefs that the agent holds independently of other beliefs. The total set of beliefs can be understood as the logical closure of the

belief base. Another early theory which focused on storing arguments is the Truth Maintenance System (Doyle, 1979). An elaborate new development in the same general category is Tennant's dependency network approach (Tennant, 2012).

However important as a research program in its own right, finding ways to represent the argumentative structure of a belief system is still not accounting for argumentation, i.e., the process of giving and taking arguments. Argumentation, in this sense, is a social practice involving more than one participant. Understanding argumentation in a belief revision framework presupposes a multi-agent, or social network, approach to belief revision. One promising point of departure for a possible marriage between belief revision and argumentation is the influential view on argumentation presented in Dung (1995). Dung is concerned with the acceptability of arguments, the fundamental principle being that the one who has the last word laughs best. Dung gives the following examples:

Mary: My government cannot negotiate with your government because your government doesnt even recognize my government.

John: Your government doesn't recognize my government either.

Mary: But your government is a terrorist government.

Assuming (somewhat unrealistically) that the debate stopped there, Mary won the argument since she had the last word. The general idea is that a statement is believable if it can be argued successfully against attacking

arguments, and so whether or not a rational agent believes in a statement depends on whether or not the argument supporting this statement can be successfully defended against the counterarguments (Dung, 1995, p. 323). It is easy to imagine how this model could be used to study belief revision as arising from the dynamics of argumentation.

Formally, an argumentation framework in Dung's sense is a pair $\langle$AR, attacks$\rangle$, where AR is a set of arguments and attacks is a binary relation on AR. In Dung's framework an argument is viewed as an abstract entity whose role is determined by its relation to other arguments. No special attention is paid to the internal structure of arguments. We will follow Dung in this respect. However, we will focus on a type of argumentation that does not immediate fit into the last word wins category. Consider the following debate on the future of the euro:

Mary: I believe the Eurozone will survive because the German Bundesbank has strongly committed itself the future of the euro.

John: I doubt you are right considering the strong political divisions that exist within the Eurozone.

Mary: I grant this, but it should also be pointed out that in the most recent meeting among the Eurozone leaders a strong statement was issued in favor of increased political integration and unity.

Here we wouldn't say that Mary won the argument in the strict sense, although she did have the last word. Rather, both parties have contributed to

the argumentation process by putting various pro and con considerations on the table, all of which deserve to be taken into account and none of which suffices by itself to establish anything beyond reasonable doubt.

What we will be concerned with, then, is not so much the strict acceptability of an argument as its more general persuasive effect. This effect has been studied extensively in social psychology within the influential and empirically robust Persuasive Argument Theory (PAT) tradition (e.g. Vinokur and Burnstein (1978), Isenberg (1986)). According to PAT, an individual's position on an issue is a function of the number and persuasiveness of pro and con arguments that the person recalls from memory when formulating his or her own position. Thus in assessing the guilt or innocence of an accused in trial, jurors come to predeliberation decisions on the basis of the relative number and persuasiveness of arguments favoring guilt or innocence. Argumentation, or group deliberation as we will also call it, will cause an individual to shift in a given direction to the extent that the discussion exposes that individual to persuasive arguments favoring that direction rather than to arguments favoring the opposite direction. How persuasive an argument is to a given individual is determined by the validity and novelty of the argument. One factor, among several, affecting perceived validity is the extent to which the argument fits into the person's previous views. Novelty has to do with how new and unusual the argument is to the person in question. Everything else equal, a novel argument has a greater persuasive force than a common place argument.

Our model, which we call Laputa (e.g. (Angere, forthcoming), (Olsson,

2011), (Vallinder and Olsson, 2012)), is similar in spirit to PAT. Laputa is also based on the assumption that the persuasive effect of an argument depends essentially on two factors: its perceived validity (including the trustworthiness of the presenter) and novelty. There are some differences, though. For instance, Laputa is more specific than PAT in assuming that individual inquirers update their degrees of belief in a particular way; namely, that dictated by Bayesianism. PAT, as such, does not postulate any more specific updating mechanism, let alone a Bayesian one. Laputa assumes, in addition, that individuals' degrees of trust are dynamically updated in a Bayesian fashion.

Furthermore, inquirers in Laputa engaging in group deliberation update their credences in a piecemeal, or sequential, fashion. The presentation of a novel argument, or collection of arguments, will normally affect the receiving inquirer's credence in the conclusion. As PAT is normally formulated, inquirers are supposed to collect in memory all the arguments they are presented with during group deliberation, postponing their own verdict on the matter until deliberation has come to an end. When the deliberation has ended the inquirer takes a stand on the basis of a holistic assessment of the number and merits of the pro and con arguments retained in memory. This holistic aspect of PAT is not unproblematic in the light of experiments indicating that the order in which arguments are presented will affect the conclusion reached. Thus, Kaplan (1977) found that subjects tend to recall persuasive arguments that they had been exposed to most recently rather than the ones they had been exposed to first.

The paper shall proceed by first informally introducing Laputa as a theory of argumentation in a group. Subsequently, the dynamical formula of Laputa are derived from accepted Bayesian principles. The paper concludes by briefly describing how this theoretical framework has been implemented in a computer program baring the same name, followed by a very brief summary of its applications to date.

2 An informal introduction to Laputa

Traditional belief revision has focused on the single inquirer setting. We wish to study belief updating in a social (network) context. An interesting complication is that inquirers in a social network not only update their degrees of belief but also their degrees of trust in their interlocutors. Ideally, we would like to have a model featuring both a rich language and a rich cognitive state representation and dynamics. However, as the matter is already quite complex— especially the proper handling of trust—some sort of compromise is necessary at the present state of investigation. One compromise would involve having a rich language but simplifying the state representation and dynamics. We will choose the opposite strategy by adopting a simple language but a rich cognitive state representation and dynamics.

In fact, the Laputa model has an extremely simple language consisting of only two propositions: p, not-p. The proposition p can stand for The eurozone will disintegrate in 2012 or John was at the party last night or anything else with a truth value. Thus p and not-p can be seen as the two

potential answers to the socially debated question: Whether p is the case?. While the language is simple, the cognitive state representation and dynamics will be quite complex. Our formal framework of choice will be Bayesianism, where belief states are represented as credences/subjective-probabilities and the basic method of belief updating is conditionalization. A social network is conceived as a set of inquirers with links between them. If there is a link from inquirer A to inquirer B that means that A can send a message to B. All inquirers focus on answering the question whether p is true. Each inquirer assigns to p at time t a certain credence, $C^t(p)$ (subjective probability). The messages inquirers can send are either "p" or "not-p". The preferred interpretation is the following:

- "σ sends the message p" means "σ gives a reason/argument for p"
- "σ sends the message not-p" means "σ gives a reason/argument against p"

Under this interpretation, the model is a model of deliberation/argumentation where reasons/arguments are, as in the Dung model, treated as black boxes. Inquirers also have a private signal they can listen to representing contributions to the deliberative process from external sources. In addition, each inquirer assigns to each information source a certain trust at t.

Now there are two main problems that need to be solved in order to make this model work: The Credence Problem and The Trust Problem. The former concerns how to update an inquirer's credence in p given new information, while the latter concerns how to update an inquirer's trust in a given source in

responce to information from that source. Being good Bayesians, we want to solve these two problems by means of conditionalization on new evidence. For the credence problem this means computing $C^{t+1}(p) = C^t(p|\sigma \text{ says that } p)$ or $C^{t+1}(p) = C^t(p|\sigma \text{ says that not-}p)$ depending on the incoming evidence. But how do we compute the right hand side of these equations? Our new credence in p after having listened to σ will depend on how much trust we placed in σ. Hence, already the credence problem requires that we also model epistemic trust; but how?

The proposal is that trust is also a form of credence; namely, a credence in the reliability of the source. This idea goes back to the Scandinavian School of Evidentiary Value (e.g. Hansson (1983), and it has been used extensively in the literature on epistemic coherence (e.g. Bovens and Olsson (2000), Olsson (2002) and Olsson (2005)).

Definition: By a source σ's degree of reliability with respect to p we shall mean the (objective) probability that σ says that p given (i) that σ says anything at all on p and (ii) that p is true.

In modeling trust Laputa takes into account every possible degree of reliability and every possible degree of unreliabilityby assigning a credence to the proposition expressing that the source is reliable/unreliable to that degree. For example, an inquirer's trust function assigns a certain credence to the proposition that the source is 75 percent reliable. Thus, trust values are here seen as second order probabilities: subjective probabilities about objective probabilities.

Returning to the credence problem, our strategy will be to proceed in two steps:

Step 1: Addressing the credence problem for one source.

Step 2: Extending this solution to a solution to the credence problem for n sources.

We need a few more assumptions before any useful work can be done:

Source Symmetry: σ's reliability with respect to p equals σ's reliability with respect to not-p.

Principal Principle: An inquirer α assigns credence ρ to the proposition that σ will report that p on the assumptions that (i) σ reports anything at all (ii) p is true and (iii) σ is reliable to degree ρ[1].

Communication Independence: Whether a source σ says something at all is independent of whether p is true as well as of σ's degree of reliability.

Once these assumptions are in place, the rest is pure mathematics, as we shall see in the next section. What about the case of n sources? Here the new credence would be calculated as $C^{t+1}(p) = C^t(p|\sigma_1 \text{ says } p,\ \sigma_2 \text{ says not-}p, ...)$. To facilitate the calculations of the new credence in this case, we need to make a further assumption:

[1] A question mark about whether this use of the Principal Principle is valid in a context where α's total evidence includes that σ has reported p has been raised (see Meacham (2010, p. 411-413)). Here, the principle is assumed as an expediency to be justified empirically on the basis of how well Laputa models deliberation and debate.

Source Independence: Inquirers take information they receive from other sources to be independent evidence.

Source Independence can be expressed in a standard way as a form of conditional independence: The credence assigned to the proposition that source σ_1 will report that p is independent of the credence assigned to the proposition that source σ_2 will report that p, and so on, conditional on the truth/falsity of p. This assumption is often used in the literature on epistemic coherence and in artificial intelligence (e.g. Pearl (1988)). As we shall see, given Source Independence the general credence problem has a simple mathematical solution.

Let us finally now turn to the trust problem: the problem of stating how to update an inquirer's trust function in the light of new evidence. Fortunately, no additional assumptions are needed to solve the trust problem (and we don't need Source Independence). In the case where the source says that p:

$$\begin{aligned}\text{Trust in } \sigma \text{ as a source on } p &= C_\alpha^{t+1}(\sigma \text{ is reliable to degree } \rho) \\ &= C_\alpha^t(\sigma \text{ is reliable to degree } \rho | \sigma \text{ says that } p),\end{aligned}$$

where the latter is a function of (i) ρ,(ii) $C_\alpha^t(p)$, and (iii) the inquirer's trust function for σ at t (or rather its expected value). Now that we have developed a sense of how this works, let us move on to the formal details.

3 Derivation of the Update Functions

This section follows the structure of (Angere, forthcoming), though it deviates from his presentation in its explicit incorporation of the present interpretation of efficacious testimony as the receipt of a novel argument/reason for, or against, p. In Bayesian fashion, the epistemic state of an individual α at time t is given by a *credence function* $C_\alpha^t : \mathcal{L} \to [0, 1]$, where we can take $\mathcal{L}$ to be a classical propositional language. The expression $C_\alpha^t(p) = x$ should be read as "Agent α's credence in proposition p at time t is x.". Let t be the time just prior to the first round of debate, $t + 1$ be the time just prior to the 2nd, and so on up to $t + n$; the time just after deliberation concludes. Laputa works by determining the value of $C_\alpha^{t+1}(p)$ from C_α^t on the basis of any novel argument α receives, or private inquiry α makes, in the period t to $t + 1$ for all α partaking in the debate, then subsequently determining the value of $C_\alpha^{t+2}(p)$ from C_α^{t+1} for all α on the same basis in the period $t + 1$ to $t + 2$, and so on all the way up to the determination of $C_\alpha^{t+n}(p)$ for all α. This sequential nature of such updates is why Laputa simulates, rather than models, debates.

Bayesian epistemology includes a principle—the principle of conditionalization—which allows $C_\alpha^{t+j}(p)$ to be determined from C_α^{t+i}, where $j = i + 1$. Let E_α^{t+i} represent all inquiry conducted, and novel arguments received, by α in the period $t+i$ to $t+j$ regarding p, then the principle of conditionalization states that $C_\alpha^{t+j}(p) = C_\alpha^{t+i}(p|E_\alpha^{t+i})$. The problem becomes one of finding an expression for $C_\alpha^{t+i}(p|E_\alpha^{t+i})$ that allows its value to be computed from information

available at $t + i$.

Relative to some inquirer α and time t, other inquirers in the network can be considered as *sources* of novel (for α at t) arguments for, or against, p. Let S be an argument either for, or against, p, let $S_{\sigma\alpha}$ be the proposition 'S is received by α from σ', let S^t_α be the proposition 'S is novel for α at t and received by α in the period t to $t + 1$'. Let $S^t_{\sigma\alpha}$ be the proposition $S_{\sigma\alpha} \wedge S^t_\alpha$ and let S^+ and S^- be the propositions 'S is for p.' and 'S is against p.', respectively. Note also that an argument against p is an argument for not-p.

But what of each inquirer's private inquiry? It is not reasonable to interpret such inquiry as the receipt of an argument. For instance, how can a first person empirical investigation result in receipt of an argument for p? What such investigations can give w.r.t. p are novel *reasons* to believe p. All arguments for p are reasons to believe p, but not all reasons to believe p are arguments. Let i be the "own inquiry source" of reasons to believe p such that each α participant has such a source. Let $S_{i\alpha}$ be the proposition 'S is received by α from i'. Where this proposition holds, then S is to be interpreted as a reason to believe either p or $\neg p$, let S^t_α be the proposition 'S is novel for α at t and received by α in the period t to $t + 1$'. Let $S^t_{i\alpha}$ be the proposition $S_{i\alpha} \wedge S^t_\alpha$.

We can then define the *reliability* of σ as a source of arguments on p for α as:

$$R^{\pm}_{\sigma\alpha} =_{\text{df.}} \mathrm{P}(S^+|S_{\sigma\alpha} \wedge p) = \mathrm{P}(S^-|S_{\sigma\alpha} \wedge \neg p),$$

where $\mathrm{P}(S^+ \mid S_{\sigma\alpha} \wedge p)$ is the objective probability with which a p-argument

is an argument for p, when α receives that argument from σ and p is true and similarly for $\mathrm{P}(S^- \mid S_{\sigma\alpha} \wedge \neg p)$. This probability might be interpreted as a (conceptual) relative frequency, or as a propensity. It is also useful to define

$$R^{\mp}_{\sigma\alpha} =_{\text{df.}} \mathrm{P}(S^+|S_{\sigma\alpha} \wedge \neg p) = \mathrm{P}(S^-|S_{\sigma\alpha} \wedge p)$$

Plainly, $R^{\pm}_{\sigma\alpha} = 1 - R^{\mp}_{\sigma\alpha}$ and note that the reliability of a source for an inquirer is assumed to be constant through time.

Similarly, we can define the reliability of an agent's own inquiry as:

$$R^{\pm}_{i\alpha} =_{\text{df.}} \mathrm{P}(S^+|S_{i\alpha} \wedge p) = \mathrm{P}(S^-|S_{i\alpha} \wedge \neg p),$$

where $\mathrm{P}(S^+ \mid S_{i\alpha} \wedge p)$ is the objective probability with which a p-reason is a reason to believe p, when α receives that reason from their own inquiry and p is true. Again, it is also useful to define

$$R^{\mp}_{i\alpha} =_{\text{df.}} \mathrm{P}(S^+|S_{i\alpha} \wedge \neg p) = \mathrm{P}(S^-|S_{i\alpha} \wedge p).$$

Now we assume that at any time t, α has a credence distribution over the reliability of σ as a source of novel arguments/reasons for them on p. Because such reliabilities can take any real value between 0 and 1, this distribution must be represented by a continuous density function ${\tau^t_{\sigma\alpha}}^2$. The credence at

[2]The derivation proceeds on this rigorous basis, but in fact the computer program Laputa approximates continuous $\tau^t_{\sigma\alpha}$ by a discrete distribution over the following set of values for $R^{\pm}_{\sigma\alpha}$: $\{0, \frac{1}{40}, \frac{2}{40}, \ldots, 1\}$. Laputa does this to make evaluation of the integrals needed to calculate the expected reliabilities—which are what is required for updating credence and trust (see later)—computationally tractable. It was found by trial and error

t that α has that the reliability of σ for them lies in the interval $[a, b]$ is given by the integral of $\tau^{t+}_{\sigma\alpha}$ between these limits:

$$C^t_\alpha(R^{\pm}_{\sigma\alpha} \in [a, b]) = \int_a^b \tau^{t+}_{\sigma\alpha}(\rho)\, d\rho.$$

Given α's credence distribution over σ's reliability we can also determine the reliability of σ that α should expect. Let $\langle \tau^{t+}_{\sigma\alpha} \rangle$ be this expected value, then

$$\langle \tau^{t+}_{\sigma\alpha} \rangle = \int_0^1 \rho \tau^{t+}_{\sigma\alpha}(\rho)\, d\rho.$$

It is also useful to define $\tau^{t-}_{\sigma\alpha}$ so that:

$$C^t_\alpha(R^{\mp}_{\sigma\alpha} \in [a, b]) = \int_a^b \tau^{t-}_{\sigma\alpha}(\rho) d\rho = \int_{1-b}^{1-a} \tau^{t+}_{\sigma\alpha}(\rho) d\rho = C^t_\alpha(R^{\pm}_{\sigma\alpha} \in [1-b, 1-a]),$$

and consequently, that the expected value of $R^{\mp}_{\sigma\alpha}$ is $\int_0^1 \rho \tau^{t-}_{\sigma\alpha}(\rho)\, d\rho = \langle \tau^{t-}_{\sigma\alpha} \rangle = 1 - \langle \tau^{t+}_{\sigma\alpha} \rangle$. In Laputa, $\tau^{t+}_{\sigma\alpha}$ is referred to as α's *trust* (in σ at t) function. The extent to which trust understood in this way corresponds to what we typically mean by trusting a source is debatable, but it does seem to capture at least part of what trust in a source is about.

that 40-step discrete distribution offered the best balance between accuracy and required computing time.

3.1 The credence update function

Now we consider the effect on α's credence of receiving a positive novel argument/reason from a single source σ. By conditionalization we have:

$$C_\alpha^{t+1}(p) = C_\alpha^t(p|S_{\sigma\alpha}^t \wedge S^+).$$

In words, α's credence at $t+1$ in p equals α's credence at t in p, given that α has received a novel argument/reason S on p from σ in the period t to $t+1$ and S is for p. Similarly, conditionalization gives $C_\alpha^{t+1}(p) = C_\alpha^t(p|S_{\sigma\alpha}^t \wedge S^-)$, so we have an expression that allows us to calculate the effect of α receiving a novel argument/reason from σ on α's credence in p whether that argument/reason is for (S^+), or against (S^-), p.

$$C_\alpha^{t+1}(p) = C_\alpha^t(p|S_{\sigma\alpha}^t \wedge S^\pm)$$

By Bayes theorem and the theorem of total probability, this gives:

$$C_\alpha^{t+1}(p^+) = \frac{C_\alpha^t(p^+)C_\alpha^t(S_{\sigma\alpha}^t \wedge S^\pm|p^+)}{C_\alpha^t(p^+)C_\alpha^t(S_{\sigma\alpha}^t \wedge S^\pm|p^+) + C_\alpha^t(p^-)C_\alpha^t(S_{\sigma\alpha}^t \wedge S^\mp|p^-)}, \quad (1)$$

where p^+ is p and p^- is not-p. $C_\alpha^t(p^+)$ is given and $C_\alpha^t(p^-) = 1 - C_\alpha^t(p^+)$, so $C_\alpha^{t+1}(p^+)$ is assuredly determined in the model if values for $C_\alpha^t(S_{\sigma\alpha}^t \wedge S^+|p^+)$, $C_\alpha^t(S_{\sigma\alpha}^t \wedge S^-|p^+)$, $C_\alpha^t(S_{\sigma\alpha}^t \wedge S^-|p^-)$ and $C_\alpha^t(S_{\sigma\alpha}^t \wedge S^+|p^-)$ are determined in the model. Each of these credences can be expanded using the

continuous version of the conditional total probability theorem.

$$C^t_\alpha(S^t_{\sigma\alpha} \wedge S^{\pm}|p^{\pm}) = \int_0^1 C^t_\alpha(S^t_{\sigma\alpha} \wedge S^{\pm}|R^{\pm}_{\sigma\alpha} = \rho \wedge p^{\pm}) C^t_\alpha(R^{\pm}_{\sigma\alpha} = \rho|p^{\pm}) d\rho$$
$$C^t_\alpha(S^t_{\sigma\alpha} \wedge S^{\mp}|p^{\pm}) = \int_0^1 C^t_\alpha(S^t_{\sigma\alpha} \wedge S^{\mp}|R^{\mp}_{\sigma\alpha} = \rho \wedge p^{\pm}) C^t_\alpha(R^{\mp}_{\sigma\alpha} = \rho|p^{\pm}) d\rho$$

Laputa then assumes that in every C and for all ρ; $S^t_{\sigma\alpha}$, $R^{\pm}_{\sigma\alpha} = \rho$, $R^{\mp}_{\sigma\alpha} = \rho$ and p are independent of each other (the communication independence assumption). This allows the above to be manipulated into the following forms by using the definition of conditional probability, cancelling terms and noting that where $C^t_\alpha(R^{\pm}_{\sigma\alpha} = \rho)$ and $C^t_\alpha(R^{\mp}_{\sigma\alpha} = \rho)$ appear in the integral they stand for the density functions $\tau^{t+}_{\sigma\alpha}(\rho)$ and $\tau^{t-}_{\sigma\alpha}(\rho)$, respectively:

$$C^t_\alpha(S^t_{\sigma\alpha} \wedge S^{\pm}|p^{\pm}) = C^t_\alpha(S^t_{\sigma\alpha}) \int_0^1 C^t_\alpha(S^{\pm}|S^t_{\sigma\alpha} \wedge p^{\pm} \wedge R^{\pm}_{\sigma\alpha} = \rho) \tau^{t+}_{\sigma\alpha}(\rho) d\rho \quad (2)$$
$$C^t_\alpha(S^t_{\sigma\alpha} \wedge S^{\mp}|p^{\pm}) = C^t_\alpha(S^t_{\sigma\alpha}) \int_0^1 C^t_\alpha(S^{\mp}|S^t_{\sigma\alpha} \wedge p^{\pm} \wedge R^{\mp}_{\sigma\alpha} = \rho) \tau^{t-}_{\sigma\alpha}(\rho) d\rho. \quad (3)$$

By the aforementioned Principal Principal we have:

$$C^t_\alpha(S^{\pm}|S^t_{\sigma\alpha} \wedge p^{\pm} \wedge R^{\pm}_{\sigma\alpha} = \rho) = \rho \quad (4)$$
$$C^t_\alpha(S^{\mp}|S^t_{\sigma\alpha} \wedge p^{\pm} \wedge R^{\mp}_{\sigma\alpha} = \rho) = \rho \quad (5)$$

Using (4) and (5) to make substitutions into (2) and (3) we have:

$$C_\alpha^t(S_{\sigma\alpha}^t \wedge S^{\pm}|p^{\pm}) = C_\alpha^t(S_{\sigma\alpha}^t)\int_0^1 \rho\tau_{\sigma\alpha}^{t+}(\rho)d\rho = C_\alpha^t(S_{\sigma\alpha}^t)\langle\tau_{\sigma\alpha}^{t+}\rangle, \tag{6}$$

$$C_\alpha^t(S_{\sigma\alpha}^t \wedge S^{\mp}|p^{\pm}) = C_\alpha^t(S_{\sigma\alpha}^t)\int_0^1 \rho\tau_{\sigma\alpha}^{t-}(\rho)d\rho = C_\alpha^t(S_{\sigma\alpha}^t)\langle\tau_{\sigma\alpha}^{t-}\rangle. \tag{7}$$

Finally, substitution of (6) and (7) back into (1) gives:

$$C_\alpha^{t+1}(p^+) = C_\alpha^t(p^+|S_{\sigma\alpha}^t \wedge S^{\pm}) = \frac{C_\alpha^t(p^+)\langle\tau_{\sigma\alpha}^{t\pm}\rangle}{C_\alpha^t(p^+)\langle\tau_{\sigma\alpha}^{t\pm}\rangle + C_\alpha^t(p^-)\langle\tau_{\sigma\alpha}^{t\mp}\rangle}. \tag{8}$$

E.g. if the novel (for α at t) p-argument/reason S, received by α from σ in the period t to $t+1$ is for p, then we read the top line of plus/minus signs in (9) to give:

$$C_\alpha^{t+1}(p^+) = C_\alpha^t(p^+|S_{\sigma\alpha}^t \wedge S^{+}) = \frac{C_\alpha^t(p^+)\langle\tau_{\sigma\alpha}^{t+}\rangle}{C_\alpha^t(p^+)\langle\tau_{\sigma\alpha}^{t+}\rangle + C_\alpha^t(p^-)\langle\tau_{\sigma\alpha}^{t-}\rangle}.$$

In any period α might receive a novel argument/reason from any one of its sources. Let $\sum_\alpha^{t+}$ be the set of sources from which α receives novel arguments/reasons for p at t, $\sum_\alpha^{t-}$ be the set of sources from which α receives novel arguments/reasons for p at t and $\sum_\alpha^t = \sum_\alpha^{t+} \bigcup \sum_\alpha^{t-}$. Then, again by

conditionalization, Bayes theorem and the law of total probability, we have:

$$C_\alpha^{t+1}(p^+) = C_\alpha^{t+1}\left(p^+ | \bigwedge_{\sigma\in\sum_\alpha^t} S_{\sigma\alpha}^t \wedge S^\pm\right)$$
$$= \frac{C_\alpha^t(p^+)C_\alpha^t\left(\bigwedge_{\sigma\in\sum_\alpha^t} S_{\sigma\alpha}^t \wedge S^\pm | p^+\right)}{C_\alpha^t(p^+)C_\alpha^t\left(\bigwedge_{\sigma\in\sum_\alpha^t} S_{\sigma\alpha}^t \wedge S^\pm | p^+\right) + C_\alpha^t(p^-)C_\alpha^t\left(\bigwedge_{\sigma\in\sum_\alpha^t} S_{\sigma\alpha}^t \wedge S^\mp | p^-\right)},$$

Laputa now makes the assumption that all α's sources are independent of one another for α; hence that:

$$C_\alpha^t\left(\bigwedge_{\sigma\in\sum_\alpha^t} S_{\sigma\alpha}^t \wedge S^\pm | p^\pm\right) = \prod_{\sigma\in\sum_\alpha^t} C_\alpha^t(S_{\sigma\alpha}^t \wedge S^\pm | p^\pm) = \prod_{\sigma\in\sum_\alpha^t} C_\alpha^t(S_{\sigma\alpha}^t)\langle\tau_{\sigma\alpha}^{t+}\rangle$$
$$C_\alpha^t\left(\bigwedge_{\sigma\in\sum_\alpha^t} S_{\sigma\alpha}^t \wedge S^\mp | p^\pm\right) = \prod_{\sigma\in\sum_\alpha^t} C_\alpha^t(S_{\sigma\alpha}^t \wedge S^\mp | p^\pm) = \prod_{\sigma\in\sum_\alpha^t} C_\alpha^t(S_{\sigma\alpha}^t)\langle\tau_{\sigma\alpha}^{t-}\rangle$$

By substitution and cancellation of terms into the above this gives:

$$
\begin{aligned}
C_\alpha^{t+1}(p^+) &= C_\alpha^t\left(p^+|\bigwedge_{\sigma\in\sum_\alpha^t} S_{\sigma\alpha}^t \wedge S^\pm\right) \\
&= \frac{C_\alpha^t(p^+)\prod\limits_{\sigma\in\sum_\alpha^t} C_\alpha^t(S_{\sigma\alpha}^t)\langle\tau_{\sigma\alpha}^{t\pm}\rangle}{C_\alpha^t(p^+)\prod\limits_{\sigma\in\sum_\alpha^t} C_\alpha^t(S_{\sigma\alpha}^t)\langle\tau_{\sigma\alpha}^{t\pm}\rangle + C_\alpha^t(p^-)\prod\limits_{\sigma\in\sum_\alpha^t} C_\alpha^t(S_{\sigma\alpha}^t)\langle\tau_{\sigma\alpha}^{t\mp}\rangle} \\
&= \frac{C_\alpha^t(p^+)\prod\limits_{\sigma\in\sum_\alpha^t} C_\alpha^t(S_{\sigma\alpha}^t)\prod\limits_{\sigma\in\sum_\alpha^t}\langle\tau_{\sigma\alpha}^{t\pm}\rangle}{C_\alpha^t(p^+)\prod\limits_{\sigma\in\sum_\alpha^t} C_\alpha^t(S_{\sigma\alpha}^t)\prod\limits_{\sigma\in\sum_\alpha^t}\langle\tau_{\sigma\alpha}^{t\pm}\rangle + C_\alpha^t(p^-)\prod\limits_{\sigma\in\sum_\alpha^t} C_\alpha^t(S_{\sigma\alpha}^t)\prod\limits_{\sigma\in\sum_\alpha^t}\langle\tau_{\sigma\alpha}^{t\mp}\rangle} \\
&= \frac{C_\alpha^t(p^+)\prod\limits_{\sigma\in\sum_\alpha^t}\langle\tau_{\sigma\alpha}^{t\pm}\rangle}{C_\alpha^t(p^+)\prod\limits_{\sigma\in\sum_\alpha^t}\langle\tau_{\sigma\alpha}^{t\pm}\rangle + C_\alpha^t(p^-)\prod\limits_{\sigma\in\sum_\alpha^t}\langle\tau_{\sigma\alpha}^{t\mp}\rangle}
\end{aligned}
$$

$$
C_\alpha^{t+1}(p^+) = \frac{C_\alpha^t(p^+)\prod\limits_{\sigma\in\sum_\alpha^{t+}}\langle\tau_{\sigma\alpha}^{t+}\rangle\prod\limits_{\sigma\in\sum_\alpha^{t-}}\langle\tau_{\sigma\alpha}^{t-}\rangle}{C_\alpha^t(p^+)\prod\limits_{\sigma\in\sum_\alpha^{t+}}\langle\tau_{\sigma\alpha}^{t+}\rangle\prod\limits_{\sigma\in\sum_\alpha^{t-}}\langle\tau_{\sigma\alpha}^{t-}\rangle + C_\alpha^t(p^-)\prod\limits_{\sigma\in\sum_\alpha^{t+}}\langle\tau_{\sigma\alpha}^{t-}\rangle\prod\limits_{\sigma\in\sum_\alpha^{t-}}\langle\tau_{\sigma\alpha}^{t+}\rangle}. \tag{9}
$$

As $C_\alpha^t(p)$, $C_\alpha^t(\neg p)$, $\langle\tau_{\sigma\alpha}^{t+}\rangle$ and $\langle\tau_{\sigma\alpha}^{t+}\rangle$ are known quantities for all σ and α at t, $C_\alpha^{t+1}(p)$ can be calculated for all α's in the network of concern on the basis of a record of the novel arguments/reasons they received from their sources and whether these were for, or against, p in the chosen period.

3.2 The trust update function

$\alpha's$ trust at t in σ—α's credence distribution at t over the reliability of σ as a source of novel arguments/reasons—also calls for updating. Let

$\tau_{\sigma\alpha}^{t+1(\pm)}$ be α's credence distribution over σ's reliability as a source of novel p-arguments/reasons after receiving an argument/reason for ($\tau_{\sigma\alpha}^{t+1(+)}$), or against ($\tau_{\sigma\alpha}^{t+1(-)}$) p in the period t to $t+1$ from σ. Then by the principle of conditionalization we have:

$$\tau_{\sigma\alpha}^{t+1(\pm)}(\rho) = \{C_\alpha^{t+1}(R_{\sigma\alpha}^{\pm} = \rho) : \rho \in [0,1]\} = \{C_\alpha^t(R_{\sigma\alpha}^{\pm} = \rho | S_{\sigma\alpha}^t \wedge S^{\pm}) : \rho \in [0,1]\} \tag{10}$$

Then for each ρ we have by the definition of conditional probability and the expansion theorem:

$$\begin{aligned} C_\alpha^t(R_{\sigma\alpha}^{\pm} = \rho | S_{\sigma\alpha}^t \wedge S^{\pm}) = \\ \left(\frac{1}{C_\alpha^t(S_{\sigma\alpha}^t \wedge S^{\pm})}\right) & C_\alpha^t(S^{\pm} | R_{\sigma\alpha}^{\pm} = \rho \wedge S_{\sigma\alpha}^t \wedge p^{\pm}) C_\alpha^t(R_{\sigma\alpha}^{\pm} = \rho \wedge S_{\sigma\alpha}^t \wedge p^{\pm}) \\ & + C_\alpha^t(S^{\pm} | R_{\sigma\alpha}^{\pm} = \rho \wedge S_{\sigma\alpha}^t \wedge p^{\mp}) C_\alpha^t(R_{\sigma\alpha}^{\pm} = \rho \wedge S_{\sigma\alpha}^t \wedge p^{\mp}) \end{aligned}$$

Using (4) and (5), together with the fact that $C_\alpha^t(S^{\pm} | R_{\sigma\alpha}^{\pm} = \rho \wedge S_{\sigma\alpha}^t \wedge p^{\mp}) = 1 - C_\alpha^t(S^{\mp} | R_{\sigma\alpha}^{\pm} = \rho \wedge S_{\sigma\alpha}^t \wedge p^{\mp})$, and the independence in C_α^t of $R_{\sigma\alpha}^{\pm} = \rho$, $S_{\sigma\alpha}^t$ and p (communication independence again), we get:

$$\begin{aligned} C_\alpha^t(R_{\sigma\alpha}^{\pm} = \rho | S_{\sigma\alpha}^t \wedge S^{\pm}) &= C_\alpha^t(R_{\sigma\alpha}^{\pm} = \rho) C_\alpha^t(S_{\sigma\alpha}^t) \frac{\rho C_\alpha^t(p^{\pm}) + (1-\rho) C_\alpha^t(p^{\mp})}{C_\alpha^t(S_{\sigma\alpha}^t \wedge S^{\pm})} \\ &= C_\alpha^t(R_{\sigma\alpha}^{\pm} = \rho) \frac{\rho C_\alpha^t(p^{\pm}) + (1-\rho) C_\alpha^t(p^{\mp})}{C_\alpha^t(S^{\pm} | S_{\sigma\alpha}^t)} \end{aligned}$$

Substituted back into (10) this gives:

$$\tau_{\sigma\alpha}^{t+1(\pm)}(\rho) = \{C_\alpha^t(R_{\sigma\alpha}^\pm = \rho | S_{\sigma\alpha}^t \wedge S^\pm) : \rho \in [0,1]\} \tag{11}$$

$$= \left\{ C_\alpha^t(R_{\sigma\alpha}^\pm = \rho) \frac{\rho C_\alpha^t(p^\pm) + (1-\rho) C_\alpha^t(p^\mp)}{C_\alpha^t(S^\pm | S_{\sigma\alpha}^t)} : \rho \in [0,1] \right\} \tag{12}$$

$$= \tau_{\sigma\alpha}^{t+}(\rho) \frac{\rho C_\alpha^t(p^\pm) + (1-\rho) C_\alpha^t(p^\mp)}{C_\alpha^t(S^\pm | S_{\sigma\alpha}^t)} \tag{13}$$

What is the denominator? By applying the continuous version of the conditional expansion theorem, followed by the discrete conditional expansion theorem we have:

$$C_\alpha^t(S^\pm | S_{\sigma\alpha}^t) = \int_0^1 C_\alpha^t(S^\pm | R_{\sigma\alpha}^\pm = \rho \wedge S_{\sigma\alpha}^t \wedge p^\pm) C_\alpha^t(R_{\sigma\alpha}^\pm = \rho \wedge p^\pm | S_{\sigma\alpha}^t)$$
$$+ C_\alpha^t(S^\pm | R_{\sigma\alpha}^\pm = \rho \wedge S_{\sigma\alpha}^t \wedge p^\mp) C_\alpha^t(R_{\sigma\alpha}^\pm = \rho \wedge p^\mp | S_{\sigma\alpha}^t) d\rho$$

By (4), (5) and aforementioned independence assumptions this gives:

$$
\begin{aligned}
C^t_\alpha(S^\pm|S^t_{\sigma\alpha}) &= \int_0^1 \rho C^t_\alpha(R^\pm_{\sigma\alpha}=\rho)C^t_\alpha(p^\pm) + (1-\rho)C^t_\alpha(R^\pm_{\sigma\alpha}=\rho)C^t_\alpha(p^\mp)d\rho \\
&= C^t_\alpha(p^\pm)\int_0^1 \rho\tau^{t+}_{\sigma\alpha}(\rho)d\rho + C^t_\alpha(p^\mp)\int_0^1 (1-\rho)\tau^{t+}_{\sigma\alpha}(\rho)d\rho \\
&= C^t_\alpha(p^\pm)\int_0^1 \rho\tau^{t+}_{\sigma\alpha}(\rho)d\rho + C^t_\alpha(p^\mp)\left(\int_0^1 \tau^{t+}_{\sigma\alpha}(\rho)d\rho - \int_0^1 \rho\tau^{t+}_{\sigma\alpha}(\rho)d\rho\right) \\
&= C^t_\alpha(p^\pm)\langle\tau^{t+}_{\sigma\alpha}\rangle + C^t_\alpha(p^\mp)\left(1-\langle\tau^{t+}_{\sigma\alpha}\rangle\right) \\
&= C^t_\alpha(p^\pm)\langle\tau^{t+}_{\sigma\alpha}\rangle + C^t_\alpha(p^\mp)\langle\tau^{t-}_{\sigma\alpha}\rangle
\end{aligned}
$$

Substituting back into (13) gives the trust update function:

$$
\tau^{t+1(\pm)}_{\sigma\alpha}(\rho) = \tau^{t+}_{\sigma\alpha}(\rho)\frac{\rho C^t_\alpha(p^\pm) + (1-\rho)C^t_\alpha(p^\mp)}{C^t_\alpha(p^\pm)\langle\tau^{t+}_{\sigma\alpha}\rangle + C^t_\alpha(p^\mp)\langle\tau^{t-}_{\sigma\alpha}\rangle} \tag{14}
$$

Using equations (9) and (14), Laputa can calculate the credence function for each debate participant at $t+1$ from their immediately preceding credence functions in response to novel arguments/reasons received in the period t to $t+1$. By repeating this process for the specified number of rounds it can determine the credence distributions that results in the group from engagement in an exchange of arguments. As the update functions are complex, it helps to derive some qualitative rules for updating against which to check Laputa's performance. The qualitative update rules for credence in p are given in table 1. A '+' means that the current belief is reinforced (i.e. $C^{t+1}_\alpha(p) > C^t_\alpha(p)$ if $C^t_\alpha(p) > 0.5$, and $C^{t+1}_\alpha(p) < C^t_\alpha(p)$ if $C^t_\alpha(p) < 0.5$.), a '$-$' that the strength of the belief is weakened, and '0' that her credence is unchanged. A source

is trusted by an inquirer if its expected reliability is greater than 0.5, and a message is surprising/expected if it contains an argument/reason for something that is disbelieved/believed to some degree by the receiver. See Olsson and Vallinder (Olsson and Vallinder) for derivations.

Source trusted?	Is message surprising? No	Neither	Yes
Yes	+	+	−
Neither	0	0	0
No	−	−	+

Table 1: Qualitative rules for updating credence.

The qualitative rules for updating trust are:

Source trusted?	Is message expected? Yes	Neither	No
Yes	+	0	−
Neither	+	0	−
No	+	0	−

Table 2: Qualitative rules for updating trust.

4 Debates in Laputa

A debate in Laputa is defined by the following parameters:

A duration for the debate: The number N of time steps over which the debate occurs.

The set of participants/inquirers: $K = \{\alpha_1, \alpha_2, \ldots, \alpha_m\}$.

The set of sources for each participant: $\sum_\alpha = \{\sigma_i, \sigma_1, \ldots, \sigma_n\}$, where σ_i is their own inquiry source and $\sigma_1 - \sigma_n \in K$ are all the participants in the debate from whom they can receive arguments.

The listen chance for each of α's sources for every α in K: $P(S_{\sigma\alpha})$ is the probability that α receives a novel argument/reason from their source σ in any time step.

An assertion threshold for every α in K: T_α is to be understood as the credence each participant in the debate must have in the conclusion of an argument before they are willing to make that argument to any of their peers. A value above 0.5 indicates an agent that is only prepared to argue for what they believe, whereas a value of less than 0.5 indicates an agent that only argues against what they believe.

A reliability of personal inquiry for each α: $R^{\pm}_{\sigma_i\alpha}$.

A trust at t function for each of α's sources for each α: $\tau^{t+}_{\sigma\alpha}$

A credence in p at t for each α in K: $C^t_\alpha(p)$.

The dynamical functions constraining the step-wise evolution of the final two types of parameters are the update functions. As currently implemented Laputa assumes all the other parameters to be constant through time, though this is an assumption that could be easily relaxed in future development.

Laputa aids the user in specifying these parameters with a directed graphical interface. In this way a debate can be inputted by specifying a number

of nodes, representing debate participants, together with arrows between the nodes, depicting the source/recipient relations among the participants; the participant at the base of an arrow is a peer-source for the participant at its head. In Laputa, such a graph is called a social network. Each node and arrow then has a number of parameters that the user specifies corresponding to the above list. Each participant is defined by five parameters: initial degree of belief in p $[C^t_\alpha(p)]$, own-inquiry accuracy $[R^{\pm}_{i\alpha}]$, own-inquiry chance $[P(S_{i\alpha})]$, own-inquiry trust $[\tau^{t+}_{i\alpha}]$ and threshold of assertion $[T_\alpha]$. Likewise, each arrow from a σ_m to α is defined by two parameters: listen chance $[P(S_{\sigma_m\alpha})]$ and listen trust $[\tau^{t+}_{\sigma_m\alpha}]$.

Finally, it is of interest to study not only particular debates but the procedures such debates exemplify. In Laputa, a deliberative procedure is specified by constraints on the debate parameters listed above and attendant sampling distributions. In more detail, the topology of the deliberative procedure is specified by a set of social networks with a sampling distribution over this set, while the edge and vertex parameter sampling protocols are specified by density functions over the unit interval. These sampling distributions are supposed to be tuned to the actual frequency of the debate parameter values and topologies exhibited by the deliberative procedure. To evaluate a deliberative procedure, Laputa samples a directed graph according to the specified protocol, and then parameterises the edges and vertices according to the specified parameter sampling protocols. The result of this sampling from the constraints is the initial state of a particular debate. Laputa then simulates this debate for a specified number of steps and records the result.

Due to the stochastic nature of Laputa simulations, the same debate should be simulated a statistically significant number of times[3] and the results aggregated. This whole process is repeated a statistically significant number of times, so that a sample of debate simulations conforming to the procedure is attained. The results are then aggregated in order to evaluate the procedure as a whole.

While the sampling of edge and vertex parameters is largely unproblematic, the modeling of deliberative procedures does face significant challenges where it comes to the sampling of social networks; challenges that the Laputa research program has yet to overcome (for an introductory discussion and direction to further reading see Masterton (Masterton)). One problem is that because the edge and vertex parameters are independently and identically distributed in Laputa, isomorphic graphs model the same social network. Hence, structure constraints should be specified by a set of isomorphism classes of directed graphs with an attendent sampling distribution. However, most randomn sampling techniques work by sampling graphs and not isomorphism classes of graphs. While a sampling distribution over graphs implies a sampling distribution over isomorphism classes of graphs, one can only discern the latter distribution from the former if one sorts the graphs into their isomorphism classes. This is a non-trivial task. For instance, Nauty—the best extant graph isomorphism identifying program—takes between a thousandth and a tenth of a second to discern whether two graphs are isomorphic

[3]The computer program does not do this at present, but it is a priority for future development.

on a standard desktop depending on the number of vertices. This may sound pretty good but the typical social network constraint contains trillions of graphs, so it would take Nauty billions of years to sort the typical social network constraint into its isomorphism classes and compute the distribution over these classes implied by a random sampling of graphs. It follows that if social networks are sampled by randomly sampling directed graphs, then it is practically impossible to ascertain whether such a sampling implies the desired sampling of social networks for the model in question.

There is a way of sampling graphs that effectively samples isomorphism classes: one defines the structure constraint graph statisitically and then runs a Markov Chain Monte Carlo (MCMC) simulation to identify the Exponential Random Graph Model (ERGM) that comes closest to producing the desired distribution of graph statistics when used for sampling. As isomorphic graphs are identical in their graph statistics (degree distribution, compactness, etc), so sampling in this manner is effectively sampling isomorphism classes of graphs. This is a superior method of sampling social networks than random sampling of graphs because there is a one to one corrrespondence between isomorphism classes of graphs and social networks; however, this approach also has its limitations. One such limitation is that structurally very dissimilar graphs can have very similar graph statistics. This leads to *instability* in the MCMC simulations used to find the appropriate sampling protocol: repeated such simulations can identify distinct ERGM's capable of producing the same statistics. This leaves us with the difficult problem of deciding which of these non-equivalent models is optimal for the deliberative

procedure at hand. Another limitation is that for certain distirbutions over graph statistics such simulations may fail to find an appropriate ERGM in a reasonable timeframe.

Despite its shortcomings, the graph statistic ERGM approach sketched above is probably the best way of specifying the topology of deliberative procedures available at the present time. This is not the method of sampling social networks currently employed in Laputa. The present approach is to sample the number of vertices in the graph uniformly from a size constraint (an interval of natural numbers) and then to populate this number of vertices with edges, each of the logically possible edges having the same probability of being included in the graph. The resultant distibution heavily favours smaller networks, samples from all logically possible graphs allowed by the size constraint, and is a distirbution over graphs rather than their isomorphism classes. A future development of Laputa would be to replace this simple binomial random sampling protocol with some variant of graph statistic ERGM approach.

5 Applications and outlook

The probabilistic model we have presented is quite complex making it difficult to prove interesting analytical results. For the purposes of studying the consequences of the model, a simulation environment was created (programmer: Staffan Angere). The Laputa simulation environment allows for effortless exploration of various complex (e.g. statistical) properties of the

model (see Figure 2).

We will not describe the simulation environment here since we have done so elsewhere (see, e.g., Olsson (2011)). However it is worth noting that due to the stochastic elements in the characterization of debates—the listen chances, the reliability of individual inquiry, etc—simulations of one and the same debate may vary in outcome. Hence it is useful to distinguish between a particular debate, corresponding to a particular simulation by Laputa, and a debate-type, corresponding to a particular parameterized social network. A *deliberative procedure* can then be viewed as a set of constraints on the debate parameters; different deliberative procedures being characterized by different constraints. For instance, jury deliberation is typically undertaken by groups of between 6 and 15 in size, while the number of participants in parliamentary debates can range from the low tens all the way up toward 1000. Such constraints, together with sampling protocols, are specified in the batch window of Laputa. Laputa then generates debate-types conforming to the argumentative practice in question by sampling within these constraints according to the relevant sampling protocols. In this way Laputa can not only simulate a single debate but many such debates conforming to some deliberative procedure.

We will close this article by giving three examples of how the Laputa simulation environment has been applied in the study of statistical properties of argumentative practices.

In the introduction we encountered the influential argumentation model put forward by Dung. The aim of the model is to study the acceptability of

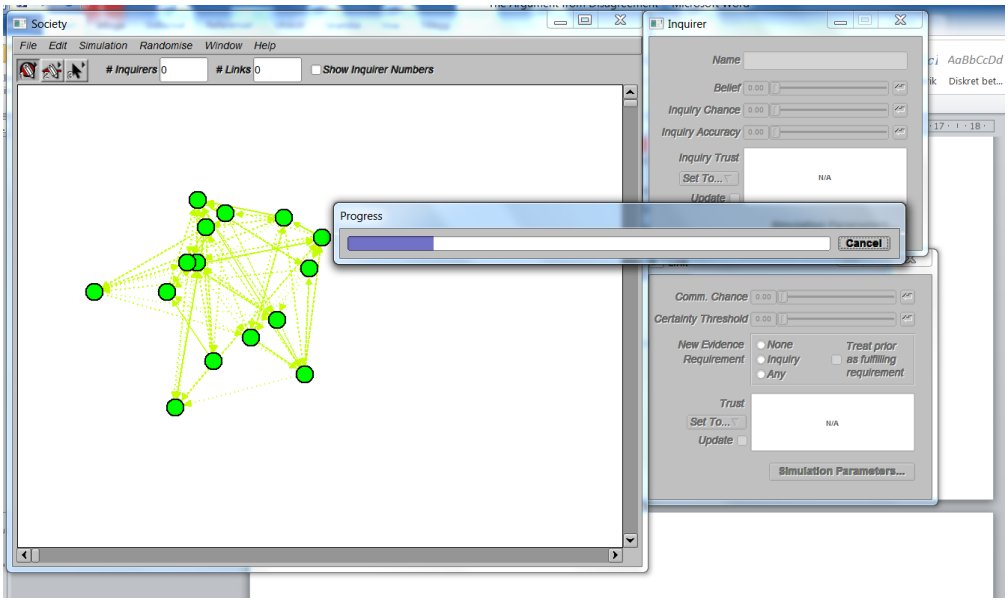

Figure 1: Laputa calculating the veritisitic value for a random sample of social networks.

arguments in a competitive, or adversarial, argumentative context in which each party wants to win over the other. We also noted that Laputa, by contrast, is a model for studying argumentation in a collaborative context where the participants put pro and con considerations on the table as part of a collective inquiry for the sake of the common good. Jury deliberation in court would be the paradigm case. Dung's model can be used, in principle, for automatic detection of the acceptability of an argument. The focus of the Laputa framework is rather on the acceptability of an argumentative practice construed as a social practice of information exchange (Goldman, 1999). More precisely, we wish to know which argumentation practices are conducive to the argumentative goals that we find valuable.

The obvious goal of an argumentation practice is to establish the truth of the proposition under consideration; e.g., the guilt of the defendant. Hence, what we are interested in, first and foremost, is the veritistic value (Goldman, 1999) of an argumentative practice, viz., an assessment of how conducive that practice is to finding the truth. But we may also be interested in the properties of argumentative practices that are not truth-related. We may, for instance, be interested in the conditions under which an argumentative practice leads to polarization whereupon members of a deliberating group predictably move toward a more extreme points in the direction indicated by the members' predeliberation tendencies (Sunstein, 2002, p. 176, italics removed). Polarization can be studied without any consideration of truth. Polarization has been observed in argumentative groups under a variety of different circumstances and is considered an empirically robust phenomenon (Isenberg, 1986).

Olsson (in press) studied the conditions under which inquirers in Laputa polarize. He found that inquirers will polarize under conditions of what he called social calibration: if initially disposed to judge along the same lines, inquirers in Laputa will adopt a more extreme position in the same direction as the effect of group deliberation, just like members of real argumentative bodies. A group is socially calibrated if either everyone correctly believes that everyone else is trustworthy or everyone correctly believes that everyone else is untrustworthy (e.g. systematic liars). Olsson noted that groups that are not socially calibrated tend to diverge in the sense that inquirers will eventually adopt contrary positions. Olsson also studied what happens to

mutual trust in the polarization process. He observed that inquirers thereby become increasingly trusting which creates a snowball effect. To the extent that Bayesian reasoning is normatively correct, the bottom line is that polarization and divergence are not necessarily the result of mere irrational group think but that even ideally rational inquirers will predictably polarize or diverge under realistic conditions.

An interesting problem in the context of collaborative inquiry concerns the conditions under which participants are warranted to make an assertion in front of their peers. This problem has a long history in philosophy. Theorists of knowledge can be divided into two camps: those who think that nothing short of certainty or (subjective) probability 1 can warrant assertion and those who disagree with this claim. Vallinder and Olsson (Vallinder and Olsson) addressed this issue by inquiring into the problem of setting the probability threshold required for assertion in such a way that the social epistemic good is maximized, where the latter is taken to be the veritistic value in the sense of Goldman (1999). Results obtained by means of computer simulation utilising Laputa indicate that the certainty rule is optimal in the (infinite) limit of inquiry and communication but that a lower threshold is preferable in less idealized cases.

Another interesting application is jury deliberation, and in particular the notorious problem of identifying the optimal size of a deliberating jury. Angere and Olsson (forthcoming) studied the effect of jury size and required majority on the quality of group decision making using an extension of the Laputa model. They found that Goldman's measure of veritistic value (Gold-

man, 1999) is unsuitable for measuring jury competence. Instead, they introduced the idea of J-value (jury value) which takes into account the unique characteristics, asymmetries and principles involved in jury voting. Using the Laputa simulation model, they found that requiring more than a 50% majority should be avoided. Moreover, while it is in principle always better to have a larger jury, given a 50% required majority, the value of having more than 12-15 jurors is likely to be negligible. Finally, they suggested a formula for calculating the optimal jury size given the cost, economic or otherwise, of adding another juror.

Laputa is more a developing framework than an already finished product. In future work, we would like to expand the analysis toolkit with a bandwagon function and other analysis tools that can help explain simulation results. In the statistical (batch) mode, Laputa generates and studies random graphs and their statistical properties. However, it is well-known that realistic social networks are often clustered in various ways. Such networks should be incorporated in future versions of Laputa. Also high on the agenda is the incorporation of division of labor and the extension of the language beyond simple yes-no questions.

References

Alchourrôn, C., P. Gärdenfors, and D. Makinson (1985). On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic 50*, 510–530.

Angere, S. Knowledge in a social network. Forthcoming.

Angere, S. and E. J. Olsson. What is the optimal size of a deliberating jury? Forthcoming.

Bovens, L. and E. J. Olsson (2000). Coherentism, reliability and Bayesian networks. *Mind 109*(436), 685–719.

Doyle, J. (1979). A truth maintenance system. *Artificial Intelligence 12*, 231–272.

Dung, P. M. (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence 77*, 321–357.

Furhmann, A. (1991). Theory contraction through base contraction. *Journal of Philosophical Logic 20*(2), 175–203.

Goldman, A. I. (1999). *Knowledge in a Social World*. Clarendon Press, Oxford.

Hansson, B. (1983). Epistemology and evidence. In P. Gärdenfors, B. Hansson, and N.-E. Sahlin (Eds.), *Evidentiary Value: Philosophical, Judicial and Psychological Aspects of a Theory*, pp. 75–97. Lund: Library of Theoria.

Hansson, S. O. (1991). Belief contraction without recovery. *Studia Logica 50*, 251–260.

Isenberg, D. (1986). Group polarisation: A critical review and meta-analysis. *Journal of Personality and Social Psychology 50*(6), 1141–1151.

Kaplan, M. F. Miller, C. E. (1977). Judgements and group discussion: Effect of presentation and memory factors on polarization. *Sociometry 40*, 337–343.

Masterton, G. Topological variability of collectives and its import for social epistemology. Forthcoming.

Meacham, C. J. G. (2010). Two mistakes regarding the principal principle. *British Journal for the Philosophy of Science 61*, 407–431.

Olsson, E. J. (2002). Corroborating testimony, probability and surprise. *British Journal for the Philosophy of Science 53*, 273–288.

Olsson, E. J. (2005). *Against Coherence:* T*ruth, Probability and Justification.* Oxford: Oxford University Press.

Olsson, E. J. (2011). A simulation approach to veritistic social epistemology. *Episteme 8*(2), 127–143.

Olsson, E. J. and A. Vallinder. Norms of assertion and communication in social networks. In press.

Pearl, J. (1988). *Probabilisitic Reasoning in Intelligent Systems.* Morgan-Kaufmann.

Sunstein, C. R. (2002). The law of group polarization. *The Journal of Political Philosophy 10*(2), 175–195.

Tennant, N. (2012). *Changes in Mind: An Essay on Rational Belief Revision.* Oxford: Oxford University Press.

Vallinder, A. and E. J. Olsson. Trust and the value of overconfidence: A Bayesian perspective on social network communication. In press.

Vallinder, A. and E. J. Olsson (2012). Does computer simulation support the argument from disagreement? *Synthese*.

Vinokur, A. and E. Burnstein (1978). Novel argumentation and attitude change: The case of polarization following group discussion. *European Journal of Social Psychology*, 335–348.