Modelling the prevalence of hidden profiles with complex argument structures

Hendrik Siebe (hendrik.siebe@gmail.com)
Department of Theoretical Philosophy, Oude Boteringestraat 52
Groningen, 9712 GL The Netherlands

Abstract

In this paper, we first introduce the 'complex hidden profile', a previously overlooked category of hidden profiles that arises from complex inferential relations among arguments. Second, in order to investigate the conditions under which interrelated arguments can generate hidden profiles, we introduce a novel Bayesian agent-based framework for collective reasoning with complex argument structures. Finally, we show that that many possible argument structures can generate hidden profiles, even when agents do not have any information in common.

Keywords: hidden profile; social deliberation; argument exchange; agent-based model; Bayesian argumentation

Introduction

Social deliberation is a form of collective reasoning and decision-making that requires its participants to exchange, weigh, and integrate reasons about how to act. An important purpose of social deliberation is to make groups more likely to identify correct decisions. It aims to achieve this by fostering a deeper collective understanding of the decision problems that a group might encounter. This could involve, for example, discussing the values and interests of relevant stakeholders, identifying the decision alternatives from which to select a course of action, and sharing arguments for or against those alternatives. Since this process takes effort and time, it is often applied to important decision problems that are too complex for any one individual to solve. However, social deliberation is not the only way in which distributed domain knowledge can be aggregated.

Even prior to social deliberation, group members typically already know information that bears on the decision problem, however incomplete or partially incorrect it may be. Assume that all the information required to find the correct answer is distributed among the group members. This information could also be combined through a voting procedure. A group of voters can collectively determine the correct answer more effectively than any of its members can, provided two conditions hold: (i) voters are more likely to be correct than not, and (ii) correlations among the votes are not too large (Dietrich & Spiekermann, 2013; Pivato, 2017). These conditions present a trade-off for social deliberation, since they are difficult to satisfy simultaneously: on the one hand group discussion might boost individual competence, but on the other hand it could introduce correlations that cause a net loss of collective competence (Dietrich & Spiekermann, 2023; Frey & van de Rijt, 2020; Siebe, 2024).

In theory, then, the ideal use case for social deliberation is one where all the necessary information to arrive at the correct answer is distributed among the group members, while none of them is more likely to be correct than not based on their private information. Unfortunately, almost 40 years of research in the tradition of the 'hidden profile' paradigm (Stasser & Titus, 1985) suggests that deliberating groups typically fail to find the correct answer under precisely these conditions. A hidden profile is a distribution of information such that the optimal alternative is hidden from each group member (Stasser & Stewart, 1992). The failure to notice and solve hidden profiles can have far-reaching consequences in a wide range of domains, such as medical diagnostics and treatment, military intelligence, business, scientific research, politics, and law (Brodbeck, Kerschreiter, Mojzisch, & Schulz-Hardt, 2007).

Thus one might ask: just how likely is a hidden profile? Numerous studies have examined the conditions that affect team performance on hidden-profile tasks (Lu, Yuan, & McLeod, 2012). However, as far as we are aware there have been no studies on the prevalence of hidden profiles in natural decision-making scenarios. In order to investigate this question, we developed a novel Bayesian agent-based model of social deliberation with complex argument structures.

In this paper we show that a significant proportion of possible argument structures can generate hidden profiles, even if all agents have unique information. These hidden profiles, induced by complex inferential relationships among arguments and henceforth referred to as 'complex hidden profiles', offer fresh insights into the hidden profile paradigm. This is because the hidden profiles from the Stasser and Titus tradition form merely a subset of the possible hidden profiles that a group might encounter. Furthermore, preliminary modeling results indicate that the well-known challenges posed by traditional hidden profiles may not generalise to complex hidden profiles. The prevalence of complex hidden profiles could thus ground a new defense of the value of social deliberation.

Traditional and complex hidden-profile tasks

In order to clarify the difference between a traditional hidden profile and what we will call a 'complex hidden profile', consider the following two scenarios.

For the first scenario, a committee is tasked with deciding whether a job applicant is suitable for the job. To this end,

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

each committee member collects evidence about the candidate. Assume that each collected piece of evidence e_i has either a positive (+1) or a negative (-1) implication for the candidate's suitability. Then, the support of a set of evidence is given by the number of pieces of pro-evidence minus the number of pieces of con-evidence.

Now it might occur that there is an *overlap* in the evidence that the committee members collect. Consider, for example, the distribution of evidence shown below in Table (1).

Table 1: Overlapping evidence generates hidden profile

person	*a*	*b*	*c*	total
pro-evidence	e_1	e_3	e_5	3
con-evidence	e_2, e_4	e_2, e_4	e_2, e_4	2
support	-1	-1	-1	+1

Since all evidence items carry the same evidential weight, the preponderance of the total collected evidence supports the applicant's suitability. However, the preponderance of each committee member's collected evidence is unfavourable to the applicant. This exemplifies the traditional hidden profile.

For the second scenario, imagine that Alice and Bob are both convinced that Charlotte would be a good supervisor for their PhD research project, but for different reasons. Alice has learned that e_1: Charlotte serves as dean of her faculty. Such a position requires good social skills and, moreover, indicates that Charlotte's colleagues entrust her with considerable responsibility. Meanwhile, Bob has found out that e_2: Charlotte is an exceptionally prolific producer of scholarly publications. This demonstrates that Charlotte is competent and intimately familiar with her field. While, considered separately, e_1 and e_2 each mark Charlotte as a desirable supervisor, their conjunction does not. Taken together, e_1 and e_2 indicate overcommitment.

This scenario is represented in Table (2) below. As before, the preponderance of the total evidence supports a different conclusion than the preponderance of each individual's accessible evidence. Yet here the result is generated by the more complex inferential structure: the conclusion is supported by different premises individually, but undermined by their conjunction. This introduces a novel category of hidden profiles that has been overlooked in previous literature.

Table 2: Inferential structure generates hidden profile

person	*a*	*b*	total
evidence	e_1	e_2	$e_1 \wedge e_2$
support	+1	+1	-1

Bayesian analysis of argumentative relations

The complex hidden profile above illustrates that the argumentative support for a conclusion need not monotonically increase in the number of supporting arguments. In order to investigate this phenomenon, we present a novel Bayesian agent-based framework. This framework can model complex argumentative structures, addressing limitations in a wide variety of prior frameworks of argument exchange, such as (Flache & Mäs, 2013; Hahn & Olsson, 2020; Taillandier, Salliou, & Thomopoulos, 2021).

Bayesian analysis

From a probabilistic perspective, the strength of an argument coincides with the evidential support that it lends to its conclusion. This idea is captured well by Bayesian approaches to argumentation, which have become relatively commonplace in argumentation theory (Verheij et al., 2016; Hahn & Hornikx, 2016).

In our Bayesian approach to arguments we treat arguments, evidence, and reasons as interchangeable entities. We abstract away from the manner of expressing arguments and evidence and focus only on whether information is communicated. During collective deliberation, even perfectly rational agents can converge on inadequate opinions if they fail to discover or share necessary evidence. Since the communication of information is at the core of the hidden profile paradigm, we represent the agents as boundedly rational epistemic agents. Specifically, our agents obey the rules of standard Bayesian learning to update their opinions as they learn more evidence. Hence, at any time, their opinions are rationally held given their evidence.

Information items are propositions that are modelled as random variables that map possible states of the world to truth values. Following classical logic, these random variables are binary, that is, they can only take one of two possible truth values, 1 (true) and 0 (false). The agents' subjective degrees of belief that a given variable is true derives from the probability distribution they assign over the corresponding states of the world. For example, our proposition might be "the die came up 6". In the absence of further information, and assuming a fair six-sided die, this proposition will be assigned the value 'true' with a probability of $\frac{1}{6}$.

Some propositions are more likely to be true (or false) given that we know the value of another proposition. For instance, if we learned that the die came up on an even number, the probability that "the die came up 6" is true equals $\frac{1}{3}$. This inferential relation between the propositions "the die came up even" and "the die came up 6" can be represented straightforwardly with a conditional probability function.

Representing inferential relations with conditional probabilities enables the non-monotonocity in defeasible inferential support that is needed to model complex hidden profiles. Because the evidential value of any particular piece of information (such as a known variable state) may be influenced by other information, we can model situations where a body of evidence opposes the conclusion that is supported by all of its components. A special case of this phenomenon is known as Simpson's paradox (Kievit, Frankenhuis, Waldorp, & Borsboom, 2013)

To illustrate how argumentative support and attack relations between propositions can be modelled with conditional probabilities, we provide a few examples. Let H be a propositional hypothesis (e.g. variable $X = 1$) to be evaluated in light of propositional evidence E (e.g. variable $Y = 0$). Then E offers 'support' for H if and only if $P(H \mid E) > P(H)$. Alternatively, E is an 'attack' on H if and only if $P(H \mid E) < P(H)$. Finally, if E neither supports nor attacks H, the two are probabilistically independent, and E is irrelevant for our evaluation of H. If there are different pieces of evidence $E_1, ..., E_k$, then the evidential force of the total evidence set towards the hypothesis can be determined by substracting $P(H)$ from $P(H \mid E_1, ..., E_k)$. A Simpson's reversal is instantiated by a set of information items that together attacks the hypothesis while each item individually supports it: $P(H \mid E_1, ..., E_k) < P(H)$ while for all $E_i \in \{E_1, ..., E_k\}$, $P(H \mid E_i) > P(H)$.

Bayesian networks

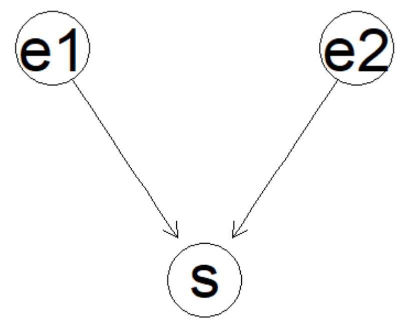
This framework sketched above can easily accommodates any set of inferential relations among propositions. Yet reasoning with chains of conditional probabilities can be aided further with Bayesian networks. A Bayesian network (BN) is a probabilistic graphical model that represents a set of random variables and their conditional dependencies via a directed acyclic graph (DAG). It is defined as a pair $< G, P >$, where G is a DAG in which the vertices (or nodes) represent random variables. In our application these variables are binary, i.e. they only have two possible states (true or false). The directed edges (or links) between the nodes represent probabilistic dependencies among the variables. If there is an edge from node A to node B, then A is called the parent of B, and B is a child of A. A child node's probability distribution is conditionally dependent on it's parent nodes' states. P is a probability function that specifies the conditional probability distribution for each variable, given its parents in the graph. Since the joint probability distribution can be unwieldy due to the exponential growth of combinations with the number of variables, Bayesian networks allow us to break it down into smaller local distributions. Each variable in the graph is associated with a conditional probability table that specifies its probability distribution conditioned on its parent variables in the DAG.[1]

Obtaining an information item (i.e. learning the realisation of a variable) can be modelled as observing the state of a propositional variable in the Bayesian network. This observation is propagated forward through the network to produce the posterior probabilities for all other variables. Figure (1) provides a simple example of a BN that models the inferential structure in Table (2). Node 's' stands for the proposition that "Charlotte is a suitable supervisor" and has two parents 'e1' and 'e2', which stand for the propositions that "Charlotte serves as dean of the faculty" and "Charlotte is an exceptionally prolific producer of scholarly publications", respectively.[2]

From the conditional probability tables, it is easy to verify that, prior to learning any information, the probability that proposition 's' is true equals $P(s = 1) = 0.42$. Once we learn that $e1 = 1$, this probability is raised to $P(s = 1 \mid e1 = 1) = 0.58$. The same posterior probability would have resulted from learning only that $e2 = 1$. However, upon learning that both $e1 = 1$ and $e2 = 1$, we find that $P(s = 1 \mid e1 = 1, e2 = 1) = 0.1$.

$P(e1=0)$	$P(e1=1)$
0.8	0.2

$P(e2=0)$	$P(e2=1)$
0.8	0.2



		$P(s=0 \mid e1,e2)$	$P(s=1 \mid e1,e2)$
$e1=0$	$e2=0$	0.7	0.3
$e1=0$	$e2=1$	0.3	0.7
$e1=1$	$e2=0$	0.3	0.7
$e1=1$	$e2=1$	0.9	0.1

Figure 1: Bayesian network model

Model

To understand how teams can run into and solve hidden-profile tasks, we have developed an agent-based model to represent a small group of individuals, the *agents*, who form *opinions* about the *target* variable that determines the correct alternative in their decision problem. The probability distribution over the target variable is unknown to the agents: their aim is to determine it on the basis of their information about the realisation of other variables. Their information items are binary *observations* of whether certain other propositions are true or not. Initially, an observation is either uniquely known by a single agent or known to all agents. When a private observation is shared, it is disclosed to all other agents as *testimony*, adding it to the public knowledge. Each agent only has an opinion about the probability that the target variable takes the value 'true', they do not have opinions about the probability that their information might be wrong or about the opinions of their peers.

[1] For more on Bayesian networks and their applications, see (Pearl, 2000)

[2] All probabilities supplied in this example are entirely fictitious and no identification with actual statistics on supervisory competence is intended. No rights can be derived from this model, nor is it meant as personal or professional advice.

The agents have a prior probability distribution over the target variable's states that is calculated from a Bayesian network that encodes the inferential structure of their collective domain knowledge. All agents access the same BN, that is, they are assumed interchangeable with respect to how any piece of evidence affects any of their beliefs. Once a new observation has been shared, they calculate their posterior probability distributions over the target node by conditionalising on both their uniquely held and the publicly known observations. Accordingly, agents who have different information items will often have different opinions about the target probability distribution. The primary outcome variables in our model are (i) the number of attempts it took to generate a hidden-profile task at the end of the first stage, and (ii) the number of correct votes at the end of social deliberation. However, for the scope of this discussion, we will focus solely on the former variable.

Model dynamics

Our agent-based model consists of a hidden-profile generation stage and a social deliberation stage. The hidden-profile task is initialised when each agent has been assigned a set of information items such that all agents form an incorrect voting preference. Subsequently, the model is updated on every time step by allowing a randomly chosen agent to share an information item with the other agents according to a predetermined strategy. Each simulation run ends when no more information can be shared.

Initialisation At the core of our model is a Bayesian network, which encodes the domain knowledge for the generated decision problem. The number of possible graphs grows exponentially in the number of nodes and each graph supports infinitely many different probability distributions. It is therefore infeasible to simulate every possible Bayesian network. Instead, we sample the space of possible Bayesian networks.

A random Bayesian network is generated in two steps. First, we generate directed acyclic graphs with a uniform probability distribution over the set of all possible graphs. Subsequently, sample from Dirichlet distributions in order to generate uniformly random (conditional) probability distributions for the nodes in the randomly generated graph.

For the target node we choose a random leaf node, that is, a node that has no 'children' (i.e. outgoing edges). For all other nodes, we either sample the Bayesian network to generate *observations* in accordance with the probabilistic structure that the BN specifies, or generate random evidence such that for each evidence node each state is equally likely to be realised. Note that the latter evidence generation method will more often generate combinations of evidence that are surprising given the probability distribution that the BN specifies. The generated observations are used to determine the posterior probability distribution over the target variable. Specifically, after conditionalising on all observations, the posterior probability that the target variable is 'true' constitutes the objective *target success probability* (TSP) that the agents will aim to discover. This success probability can be interpreted as the probability of receiving a $1 payout on a bet.

When the observations have been generated, the N agents each receive an equal number of *observations* that they hold uniquely. Additionally, it is possible to assign a number of observations to a store of *common knowledge* that all agents have access to. The agents conditionalise on their own private observations and any common knowledge to calculate their posterior *opinion* as to what the target success probability is. Based on this estimate of the TSP, they form a *voting preference* over two alternatives: to bet $0.50 (in case they estimate the success probability to be greater than $\frac{1}{2}$) or not to bet at all (otherwise).

Accordingly, there are two ways in which agents can vote incorrectly, namely, by voting to bet when the success probability is less than $\frac{1}{2}$, or by voting not to bet when the success probability is greater than $\frac{1}{2}$. Agents with different observations can form different opinions regarding the success probability, and consequently they may have different voting preferences.

A hidden profile is only generated when all agents have wrong pre-deliberative voting preferences. If there is at least one agent with a correct initial voting preference, another attempt is made to generate a hidden profile. For this next attempt, a new realisation of the evidence nodes of the same Bayesian network is generated. Then, conditional on this new combination of observations, a new objective TSP is calculated, the new observations are distributed among the agents and the common knowledge store, and the resulting voting preferences are once again checked for incorrectness. This procedure is repeated until either a hidden profile is generated, or the specified maximum number of attempts is exhausted.

Social deliberation Once a hidden profile is generated, an agent is randomly selected at each time step and given an opportunity to share one of their uniquely held observations. Shared observations are disclosed publicly, yet sharing information is assumed to be costly. Hence, the agents will only share information provided that this information (i) should be convincing to others, and (ii) has not yet been shared. What the other agents are to be convinced of depends on the selected argument exchange strategy. Our model currently implements three information exchange strategies: *random*, *advocacy*, and *conformity*. Out of these three, the advocacy and conformity strategies both model biased information sharing. Specifically, agents who adopt the advocacy strategy only share information that supports their current voting preference, while agents who communicate according to the conformity strategy only share information that supports the current majority voting preference.

Preliminary modelling results indicate that both strategies are capable of solving complex hidden profiles, although advocacy tends to be a more successful strategy than conformity. However, for our present purposes, we set aside further discussion on this stage of the model.

Implementation

The Bayesian models were constructed in the R programming language (R Core Team, 2021). The graph generation utility from the *bnlearn* package (Scutari & Denis, 2014) was used to sample uniformly from the set of possible graphs, and the *rdirichlet* function from the *MCMCpack* package (Martin, Quinn, & Park, 2011) was used to assign uniformly randomly generated condition probability distributions for the graph nodes. The belief updates via exact inference were handled by the *gRain* package (Højsgaard, 2012). The agent-based model was implemented in Netlogo v6.2 (Wilensky, 1999) and interacts with the Bayesian network via the R extension (Thiele & Grimm, 2010). Analysis of the results was carried out in R, and plots were made using the *ggplot2* library (Wickham, 2016).

Results

In this paper, we focus on the experimental outcomes obtained during the hidden profile generation stage. In order to investigate the prevalence of hidden profiles in small groups, we generated 1000 Bayesian networks by uniformly randomly sampling the space of possible BNs with 8 nodes (i.e. 7 observations). For each BN we generated 200 realisations of possible observations and corresponding posterior target success probabilities. We first looked at the proportion of BNs that generated at least one hidden profile, and then measured the overall proportion of hidden profiles across all realisations of all BNs. We tested three different conditions for how observations could be distributed among agents.

The dependent variable (DP) is the number of BNs from which a hidden profile can be constructed. The independent variable (IV) is the number of initial *positions*, where we define a position as a unique evidence base within the population of agents. We varied the IV by distributing the realisations of evidence from a BN over 3, 6, or 7 positions after the observations are distributed over the population of agents.

These conditions were set up by initialising the model with 3, 6, or 7 agents, and assigning each agent a unique set of observations. We focused on initial positions rather than population size, because the agents that occupy the same position are interchangeable, so the number of agents in the population is unimportant[3].

Table 3: Hidden profile generation difficulty

Condition	Successes	Failures	λ	SE
3 positions	968	32	26.98	0.17
6 positions	960	40	28.69	0.17
7 positions	650	350	22.73	0.19

For the first condition, we assigned 2 unique observations to each of 3 agents and assigned 1 observation to the public knowledge store. As we see in Table 3, the proportion of BNs that generated at least one hidden profile in 200 attempts is almost 97%. The second condition followed the same procedure, but distributed the observations over 6 agents and assigned 1 observation to the public knowledge store. In other words, each agent had only one uniquely held observation in addition to the single public information item. Here, the success rate for generating a hidden profile is almost 96%. For the third condition, we assigned one single unique observation to each of 7 agents, and left the store of public information items empty. Here the success rate for generating a hidden profile dropped to 65%.

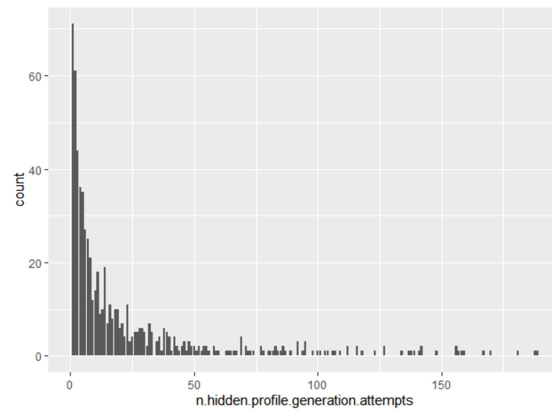


Figure 2: Counts of attempts until first hidden profile with 7 positions

The proportion of BNs that could generate at least one hidden profile differs substantially across the conditions. This is confirmed by a chi-squared test, as $\chi^2 = 544, \quad df = 2, \quad p < 2.2 \times 10^{-16}$.

Setting aside the BNs that failed to generate a hidden profile, we were interested in assessing the difficulty of the task for the successful Bayesian networks. In Figure (2) we see a typical plot of the counts of the number of networks that generated a hidden profile within 200 attempts, for 7 positions. The plots for 3 and 6 positions have the same shape. The number of attempts made before a hidden profile is generated follows a Poisson distribution. Accordingly, we fitted our data to a Poisson distribution to estimate both the mean (λ) and the standard error (SE) parameters.

Once again, the differences across the conditions are statistically significant, as is shown in Table 3. Since λ represents the average number of attempts needed for successful BNs to generate a hidden profile, a lower value indicates that it is easier to succeed. We observe that generating a hidden profile in the condition with 7 positions takes fewer attempts than in the other conditions.

We also compared the proportions of hidden profiles across all realisations of all BNs between the conditions (Table 4). With a larger number of positions, the prevalence of hidden

[3]To see why this is the case, note that we could simply multiply the number of agents by 14, 7, and 6, respectively, to normalise the population sizes across the conditions

profiles is greater despite there being fewer BNs that succeeded in generating a hidden profile at all. We also found this effect in other (small) Bayesian network sizes. Moreover, we observed that both the proportion of BNs that could generate at least one hidden profile and the proportion of hidden profiles across all the sampled realisations of all the BNs increased with the number of nodes.

Table 4: Hidden profile proportions

Condition	Successful BNs	Hidden profiles
4 nodes, 2 positions	0.259	0.010
4 nodes, 3 positions	0.208	0.025
6 nodes, 2 positions	0.662	0.026
6 nodes, 4 positions	0.651	0.033
6 nodes, 5 positions	0.453	0.045
8 nodes, 3 positions	0.968	0.053
8 nodes, 6 positions	0.960	0.060
8 nodes, 7 positions	0.650	0.071

Discussion

In the simulations of the hidden profile generation stage of our agent-based model, many randomly sampled Bayesian networks could generate a hidden profile with positive probability. We chose a network size of eight nodes to represent typical deliberation settings with a relatively small number of relevant arguments. In the conditions with common information, agents uniquely held either one or two information items while a single observation was public knowledge. The proportion of BNs that succeeded in generating at least one hidden profile under these conditions exceeded 95% for this network size. However, it is important not to conflate this success rate with the probability of encountering a hidden profile, as each Bayesian network had 200 attempts. The plot in Figure (2) shows a left-skewed distribution of successful Bayesian networks as function of the number of attempts. This indicates that the majority of successes are concentrated towards the lower end of the attempts count, with fewer occurrences at higher values. After fitting this data to a Poisson distribution to estimate the mean (λ), we obtained relatively low values of 26.98 and 28.69 for these conditions. These figures represent the average number of attempts required for a Bayesian network to generate a hidden profile, given that it could succeed at all.

In the condition where each agent possessed a single unique piece of evidence, the information was maximally fragmented. Intriguingly, we observed that this resulted in the lowest proportion of BNs that generated a hidden profile at all: only 65% of networks were successful. This might be explained by the fact that this condition generates far fewer different distributions of evidence over the agents for any realisation of any Bayesian network. However, such a fragmented distribution of evidence does not seem desirable for those seeking to evade hidden-profile tasks. Not only did it take fewer attempts ($\lambda = 22.73$) to generate a hidden profile in those BNs that were successful, even across all 1000 random BNs the proportion of generated hidden profiles was greater than in the other conditions.

What do these results mean for social deliberation? From the existing literature on hidden profiles, we already knew that common information can generate hidden profiles. One might therefore not be too surprised to learn that a large proportion of random inferential structures can generate hidden profiles with positive probability in conditions where at least one piece of evidence is held in common. Yet our simulations show not only that hidden profiles can occur even in the absence of common information, they suggest a higher overall prevalence in such situations. Moreover, a larger number relevant information items seems to increase the prevalence of hidden profiles across all conditions.

The capacity of social deliberation to facilitate beneficial information exchange has not gone unchallenged (Solomon, 2006). The notoriously poor performance of deliberating groups on traditional hidden-profile tasks is often viewed as a key argument against social deliberation (Sunstein, 2006). However, many, if not most, of the hidden profiles that occur might not be traditional hidden profiles. This opens up possibilities that have been overlooked so far within the confines of the hidden profile paradigm.

Recall the comparison between the scenarios represented in Table (1) and Table (2). Clearly, agents would be incapable of solving the traditional hidden profile presented in Table (1) if they only exchanged arguments supporting their pre-deliberative opinions. In contrast, the complex hidden profile in Table (2) readily allows this possibility.

Future work could explore the limitations of our assumptions, such as the assumptions that all agents share the exact same understanding the inferential relations that obtain among the arguments, or that they share the same interests. Similarly, the communication of arguments on different network structures remains to be explored. There is also scope for conducting more experiments with different distributions of evidence among the agents and different sizes of Bayesian networks. These are directions for future research.

Meanwhile, our agent-based model contributes to the literature on formal modelling of persuasive argument exchange in the tradition of (Burnstein & Vinokur, 1973) in an important respect: it equips Bayesian agents with the ability to reason with rich argument structures in a computationally efficient and conceptually transparent fashion.

Conclusion

While our results show that there are more hidden profiles than traditionally accounted for, these hidden profiles might be far easier for people to solve than traditional hidden profiles. This possibility sheds new light on the epistemic value of social deliberation and cautions against inferring general norms for social deliberation from the traditional hidden-profile paradigm.

Acknowledgments

I am very grateful to Jan-Willem Romeijn and Andreas Flache for their invaluable feedback. Additionally, I would like to express my deep appreciation to the anonymous reviewers for their comments on an earlier draft of this article. Many thanks go also to Freek Oude Maatman for his helpful suggestions. This study is part of the research program Sustainable Cooperation – Roadmaps to Resilient Societies (SCOOP). The author is grateful to the Netherlands Organization for Scientific Research (NWO) and the Dutch Ministry of Education, Culture and Science (OCW) for generously funding this research in the context of its 2017 Gravitation Program (grant number 024.003.025)

References

Brodbeck, F. C., Kerschreiter, R., Mojzisch, A., & Schulz-Hardt, S. (2007). Group decision making under conditions of distributed knowledge: The information asymmetries model. *The Academy of Management Review*, *32*(2), 459–479.

Burnstein, E., & Vinokur, A. (1973). Testing two classes of theories about group induced shifts in individual choice. *Journal of Experimental Social Psychology*, *9*(2), 123-137. doi: 10.1016/0022-1031(73)90004-8

Dietrich, F., & Spiekermann, K. (2013). Epistemic democracy with defensible premises. *Economics and Philosophy*, *29*(1), 87–120. doi: 10.1017/s0266267113000096

Dietrich, F., & Spiekermann, K. (2023). Jury Theorems. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Spring 2023 ed.). Metaphysics Research Lab, Stanford University. `https://plato.stanford.edu/archives/spr2023/entries/jury-theorems/`.

Flache, A., & Mäs, M. (2013). Differentiation without distancing. explaining bi-polarization of opinions without negative influence. *PLoS ONE*, *8*(11). doi: 10.1371/journal.pone.0074516

Frey, V., & van de Rijt, A. (2020). Social influence undermines the wisdom of the crowd in sequential decision making. *Management Science*, *67*(7), 4273-4286. doi: 10.1287/mnsc.2020.3713

Hahn, U., & Hornikx, J. (2016). A normative framework for argument quality: Argumentation schemes with a bayesian foundation. *Synthese*, *193*(6), 1833–1873. doi: 10.1007/s11229-015-0815-0

Hahn, U., & Olsson, J. U. (2020). Truth tracking performance of social networks: how connectivity and clustering can make groups less competent. *Synthese*, *197*, 1511–1541. doi: 10.1007/s11229-018-01936-6

Højsgaard, S. (2012). Graphical independence networks with the gRain package for R. *Journal of Statistical Software*, *46*(10), 1–26. Retrieved from `https://www.jstatsoft.org/v46/i10/` doi: 10.18637/jss.v046.i10

Kievit, R., Frankenhuis, W., Waldorp, L., & Borsboom, D. (2013). Simpson's paradox in psychological science: a practical guide. *Frontiers in Psychology*, *4*. doi: 10.3389/fpsyg.2013.00513

Lu, L., Yuan, Y. C., & McLeod, P. L. (2012). Twenty-five years of hidden profiles in group decision making: A meta-analysis. *Personality and Social Psychology Review*, *16*(1), 54-75. doi: 10.1177/1088868311417243

Martin, A. D., Quinn, K. M., & Park, J. H. (2011). Mcmcpack: Markov chain monte carlo in r. *Journal of Statistical Software*, *42*(9), 1–21. Retrieved from `https://www.jstatsoft.org/index.php/jss/article/view/v042i09` doi: 10.18637/jss.v042.i09

Mercier, H., & Sperber, D. (Eds.). (2017). *The enigma of reason*. Cambridge, MA, USA: Harvard University Press.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge University Press.

Pivato, M. (2017). Epistemic democracy with correlated voters. *Journal of Mathematical Economics*, *72*, 51-69. doi: 10.1016/j.jmateco.2017.06.001

Pollock, J. L. (1987). Defeasible reasoning. *Cognitive Science*, *11*(4), 481–518. doi: 10.1207/s15516709cog1104_4

R Core Team. (2021). *R: A language and environment for statistical computing* (Tech. Rep.). Vienna, Austria. Retrieved from `https://www.R-project.org/`

Scutari, M., & Denis, J.-B. (2014). *Bayesian networks with examples in R*. Boca Raton: Chapman and Hall.

Siebe, H. (2024). The interdependence of social deliberation and judgment aggregation. *Social Choice and Welfare*.

Solomon, M. (2006). *Groupthink* versus *the Wisdom of Crowds*: The social epistemology of deliberation and dissent. *Southern Journal of Philosophy*, *44*(S1), 28–42. doi: 10.1111/j.2041-6962.2006.tb00028.x

Stasser, G., & Stewart, D. (1992). Discovery of hidden profiles by decision-making groups: Solving a problem versus making a judgment. *Journal of Personality and Social Psychology*, *63*(3), 426-434. doi: 10.1037/0022-3514.63.3.426

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology*, *48*(6), 1467-1478. doi: 10.1037/0022-3514.48.6.1467

Sunstein, C. R. (2006). Deliberating groups versus prediction markets (or hayek's challenge to habermas). *Episteme*, *3*(3), 192–213. doi: 10.3366/epi.2006.3.3.192

Taillandier, P., Salliou, N., & Thomopoulos, R. (2021). Introducing the argumentation framework within agent-based models to better simulate agents' cognition in opinion dynamics: Application to vegetarian diet diffusion. *Journal of Artificial Societies and Social Simulation*, *24*. doi: 10.18564/jasss.4531

Thiele, J. C., & Grimm, V. (2010). Netlogo meets r: Linking agent-based models with a toolbox for their analysis. *Environmental Modelling & Software*, *25*(8), 972-974. doi:

10.1016/j.envsoft.2010.02.008

Verheij, B., Bex, F., Timmer, S. T., Vlek, C. S., Meyer, J.-J. C., Renooij, S., & Prakken, H. (2016). Arguments, scenarios and probabilities: connections between three normative frameworks for evidential reasoning. *Law, Probability and Risk*, *15*(1), 35-70. doi: 10.1093/lpr/mgv013

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from `https://ggplot2.tidyverse.org`

Wilensky, U. (1999). *Netlogo* (Tech. Rep.). Northwestern University, Evanston, IL: Center for Connected Learning and Computer-Based Modeling. Retrieved from `http://ccl.northwestern.edu/netlogo/`