Synthese (2020) 197:1203–1224
https://doi.org/10.1007/s11229-018-1743-6

The brain attics: the strategic role of memory in single and multi-agent inquiry

Emmanuel J. Genot[1] · **Justine Jacot**[1]

Received: 11 May 2017 / Accepted: 24 February 2018 / Published online: 6 March 2018

Abstract M. B. Hintikka (1939–1987) and J. Hintikka (1929–2016) claimed that their reconstruction of the 'Sherlock Holmes sense of deduction' can "serve as an explication for the link between intelligence and memory" (1983, p. 159). The claim is vindicated, first for the single-agent case, where the reconstruction captures strategies for accessing the content of a distributed and associative memory; then, for the multi-agent case, where the reconstruction captures strategies for accessing knowledge distributed in a community. Moreover, the reconstruction of the 'Sherlock Holmes sense of deduction' allows to conceptualize those strategies as belonging to a continuum of *behavioral strategies*.

Keywords Sherlock Holmes · Interrogative approach to inquiry · Associative memory · Distributed knowledge · Cognitive outsourcing

Introduction

In the 1980s, M. B. and J. Hintikka vindicated the 'Sherlock Holmes sense of deduction', that is, the notion that deduction has a strategic role in inquiry and discovery (Hintikka and Hintikka 1983, 1989). The Hintikkas further argued that their vindication was not limited to the strategic use of deduction, but could also elucidate the role of cognitive functions such as memory and attention in the process of inquiry. As an illustration, they noticed that:

✉ Emmanuel J. Genot
emmanuel.genot@fil.lu.se

Justine Jacot
justine.jacot@fil.lu.se

1 Department of Philsophy, University of Lund, Box 192, 221 00 Lund, Sweden

> [F]requently, all [Sherlock Holmes] has to do is to perform an *anamnesis* and recall certain items of information which he already had been given [...] or which are so elementary that any intelligent reader is expected to know them. (Hintikka and Hintikka 1983, p. 159)

The Hintikkas' argument goes as follows: (1) the reconstruction of the 'Sherlock Holmes sense of deduction' provides an explication of how deductive anticipations guide information-seeking through observation; (2) by the above remark, recalling information sometimes plays the same role as observation in Holmesian inquiry; (3) subsequently:

> [the] partially shared model [of the 'Sherlock Holmes sense of deduction'] for recall and intelligent inquiry may perhaps serve as an explication for the link between memory and intelligence. (Hintikka and Hintikka 1983, p. 159)

The Hintikkas phrased their conclusion cautiously, because they reconstructed only what cognitive psychology refers to as 'analytic' thought processes. In order to provide an explication for the link between intelligence and memory, their analysis would have required an extension to *associative* processes. This extension was unfortunately never realized. After the demise of M. B. Hintikka (in 1987), J. Hintikka and his collaborators developed what became the *interrogative model of inquiry* (IMI) with an exclusive focus on the relations between theorem-proving and information-seeking. In the process, the naturalistic motivations and ambitions of the original approach were almost entirely forgotten.

In this paper, we outline the extension of the Hintikkas' approach from analytic to associative processes. We vindicate the suggestion that it can illuminate "the link between intelligence and memory" by showing that memory supports *strategic associations*. The relevant concept of strategy (behavioral strategies) is borrowed from game theory, and is thus well-defined and mathematically precise. Furthermore, our appeal to this concept is grounded in computational models of human memory, rather than conceptual analysis alone. And insofar as these models are supported by empirical data, we can claim a measure of empirical adequacy for this concept of strategy. Our argument carries further a critical assessment of the IMI that we initiated with Genot (2017). We argued that the IMI suffers from shortcomings that the Hintikkas' original (naturalistic) interpretation can overcome. And we pointed out that their interpretation anticipated on empirical research that has since then clarified how the architecture of memory supports creative thinking and problem-solving.

As a tribute to the Hintikkas and Sherlock Holmes, we propose that our argument is a vindication of the *Sherlock Holmes sense of memory*. The ideas credited to Holmes about memory are grounded in ancient rhetorics, but with a modicum of interpretive charity they dovetail the description of memory from cognitive neuroscience and cognitive psychology. However, our interest lies beyond Holmes' inquiries, which primarily illustrate the single-agent case. Our reconstruction of strategic associations permits us to consider strategies for accessing information distributed across the memory of a single epistemic agent on par with strategies for accessing information distributed across a network of epistemic agents: when cooperation prevails, they belong to the same continuum.[1]

The paper builds up to the conclusion just expressed in three steps. Section 1 presents the Hintikkas' reconstruction of the strategic role of deduction (1.1), then Holmes' conception of memory in the light of cognitive neuroscience (1.2), and combines them to obtain a characterization of the strategic role of memory for the single-agent case (1.3). Section 2 introduces the multi-agent case with a non-fictional example, namely the (serendipitous) discovery of the cosmic microwave background (2.1) that we reconstruct following the Hintikkas' approach (2.2) in order to obtain a game-theoretic characterization of the information-seeking strategies it illustrates (2.3). Section 3 introduces the concept of *distributed knowledge* (3.1) and proposes that strategies for accessing information distributed in a community are continuous with strategies for accessing the distributed content of one's memory (3.2). We conclude with suggestions for further research. An appendix collects examples of deductive and non-deductive problems illustrating the relations between deduction, research questions, and recall.

Our line of argument parallels two recent trends in mainstream analytic epistemology, namely the debate on extended cognition and the burgeoning discussion of the relation between memory and distributed cognition.[2] We have however resisted the temptation to draw explicit parallels with those trends. Our reason is twofold. First, the Hintikkas' concern with the relation between memory and inquiry predates those debates by nearly thirty years. Therefore, attempting to put the former in the perspective of the latter would incur a risk of misidentification and incorrect charaterization of the Hintikkas' project. Second, the notion of 'distributed knowledge' investigated in these trends is considerably richer than the one we investigate and interpret within our relatively simple formal model. This is again intentional, since this model is intended to be as conservative an extension as possible of the Hintikkas' original model. Whether this simple model can contribute to the debate about richer notions of 'extended' and 'distributed knowledge' is therefore an open question and a topic for further research. We will suggest in conclusion some directions this research could take.

1 The Sherlock Holmes sense of memory

1.1 Memory and the 'Sherlock Holmes sense of deduction'

M. B. and J. Hintikka reconstructed the 'Sherlock Holmes sense of deduction' in inquiry with a two-player game between *Nature* and *Inquirer*. They used a deductive proof system (Beth tableaux, Beth 1955) as a bookkeeping method to keep track of Inquirer's reasoning about the underlying state of Nature, and they extended it with

[1] See Genot and Jacot (2012), Genot (2017) for a discussion of attention and memory in inquiry contexts involving interaction under less than perfect cooperation. The discussion involves examples from Sherlock Holmes lore, showing that capturing the complexity of Holmesian inquiry exceeds the resources of single-agent reconstructions.

[2] The first is extensively covered in a special issue of *Philosophical Issues* edited by A. Carter, J. Kallestrup, O. Palermos, and D. Pritchard published in 2014 (Philosophical Issues 2014). As for the second, the reader will find a comprehensive survey (as of 2013) in Michaelian and Sutton (2013).

rules allowing Inquirer to ask questions about said state of Nature. In this reconstruction, answers (when obtained) strengthen the background information available to Inquirer from the premises and warrant further conclusions about the state of Nature. In a deductive problem, one can reason to a solution through deductive inference from the premises alone. In a non-deductive problem, deductive inferences provide no solutions unless additional information is collected, 'transforming' in all effect the problem into a deductive one. The reconstruction uncovers the relations between strategies for solving deductive and non-deductive problems.

A non-deductive problem can be solved by strengthening the premises enough to warrant a deductive proof for a solution from the strengthened premises. And the best deductive strategy to derive a solution is the one that yields the simplest deductive proof. Subsequently, insights into the best *deductive* strategy for deriving a solution to a non-deductive problem from a set of premises strengthened with some information, are also insights into how to actually extend the set of premises for solving this non-deductive problem. These insights guide information-seeking, and the discovery of a solution. Furthermore, they are not the result of some irrational intuition: insofar as they are grounded in logic, they support the claim that a logic of discovery is possible.[3]

In order to make the notion of strategy formally precise, M. B. and J. Hintikka specified a utility function for an Inquirer with bounded cognitive resources [working memory (Hintikka and Hintikka 1983) and attention (Hintikka and Hintikka 1989)] and a subsequent preference for lower cognitive loads. Unfortunately, those utilities cannot in general support strategic reasoning without hypercomputation, that is, solving problems that are not computable by Turing machines [formal arguments are presented in Genot (2017), Genot and Jacot (2017)]. After the death of M. B. Hintikka, J. Hintikka and his collaborators sidestepped the issue, and bumped off utilities. Specifically, they characterized the relation between deduction and interrogation with the resources of first-order semantics and proof theory alone [see Hintikka et al. (1999) for formal results and Hintikka (2007) for informal expositions]. In a nutshell, if Inquirer has the same preferences as a theorem prover, she tries to minimize the number of cases she reasons about, and the number of individuals to consider in those cases. Expressing those preferences with utility functions becomes redundant, and the game-theoretic analysis of strategy selection is substituted with a proof-theoretic one, which is appropriately resource-sensitive.

When they substituted game theory with meta-mathematics, Hintikka and his collaborators prevented their reconstruction of interrogative inquiry to collapse into conceptual contradiction. But they also threw the proverbial baby with the bathwater, leaving implicit in their formal results any constraint that would result from Inquirer's bounded cognitive resources. We have shown in Genot (2017) that those bounds deeply impact the interpretation of the formal results expressing the relation between interrogation and deduction, in particular when instrumental questions cannot be obtained mechanically from the premises (cf. also Appendix).

[3] The notion that there can be no 'logic of discovery' has dominated philosophy of science since Popper (1992) and Reichenbach (1938). A detailed argument against this view based on the Hintikkas' analysis, with formalized examples, is presented in Genot (2017). Appendix presents semi-formalized examples of deductive problems, non-deductive problems, and deductive 'transforms' of non-deductive problems.

Of special interest are cases where the solution involves memory and attention, and hinges on the ability to recall relevant information and associate it to the premises of the problem under consideration. And in fact these cases are within the purview of the approach that M. B. and J. Hintikka had initially laid down:

> Eliciting […] information by questioning can be viewed as one possible recall procedure. At the same time, it can be generalized so as to become a common model of several different kinds of information-gathering activities, deductive as well as inductive. This partially shared model for recall and intelligent inquiry may perhaps serve as an explication for the link between memory and intelligence. For another thing, our general idea should not surprise any Sherlock Holmes fan. In some cases, the great detective has to carry out an observation or even an experiment to answer the question. More frequently, all he has to do is to perform an *anamnesis* and recall certain items of information which he already had been given and which typically had been recorded in the story or novel for the use of the readers, too, or which are so elementary that any intelligent reader is expected to know them. (Hintikka and Hintikka 1983, p. 159)

We suggested in Genot (2017) that the "link between intelligence and memory" should be construed as the ability to use memory strategically. Since memory relies on associations, the suggestion entails the existence of 'strategic associations'. This notion may seem somewhat of an oxymoron, if 'association' is understood as 'free' or 'spontaneous' but the notion has an empirically grounded interpretation. We will develop further this interpretation in Sect. 1.3 and argue that it vindicates some of Holmes' ideas on memory, to which we now turn.

1.2 Holmes on the brain-attic

Holmes' view of memory is revealed in *A Study in Scarlet*. Having casually mentioned that the Earth revolves around the Sun, Watson is first surprised to find out that Holmes ignored that fact, and then baffled when the consulting detective proclaims that he will do his best to forget everything about it. Holmes then offers the following clarification:

> I consider that a man's brain originally is like a little empty attic, and you have to stock it with such furniture as you choose. A fool takes in all the lumber of every sort that he comes across, so that the knowledge which might be useful to him gets crowded out, or at best is jumbled up with a lot of other things, so that he has a difficulty in laying his hands upon it. Now the skillful workman is very careful indeed as to what he takes into his brain-attic. He will have nothing but the tools which may help him in doing his work, but of these he has a large assortment, and all in the most perfect order. It is a mistake to think that that little room has elastic walls and can distend to any extent. Depend upon it there comes a time when for every addition of knowledge you forget something that you knew before. It is of the highest importance, therefore, not to have useless facts elbowing out the useful ones. (*A Study in Scarlet*, 2)

Holmes 'brain-attic memory' (BAM for short) builds on the *method of loci* of Greek and Roman rhetoric, expounded for instance in Cicero's *De Oratore*. The method relies on what is often referred to as 'mind palaces' in popular expositions (for instance Yates 1966), that is, visualizations of spatial structures in which items to be committed to memory are assigned specific *loci* according to a user-defined classification scheme. Holmes implicitly assumes a one-one correspondence between *loci* and actual locations in the brain, and draws the consequence that the size of the latter being bounded, some items must be 'committed out' of memory—hence his ignorance of the layout of the solar system.[4]

The BAM conception blends considerations about the functional role of memory (storage) and its physical substrate (the brain). Contemporary cognitive neuroscience depicts a more subtle relation between the functional organization of memory, and its material substrate. Memories are encoded in the brain by neurons sensitive to microfeatures that can receive a value within a range; are maximally activated in response to a particular value within that range; and respond to other values with an intensity that decreases with the distance from their preferred value (see Churchland and Sejnowski 1992). For instance, the red diamond of a card suit, depicted in a particular shade of red and with two 45° angles, activates maximally neurons that respond to that shade and that angle, less-than-maximally neurons that respond to slightly darker or lighter shades of red, and to angles of 44° and 46°, and not at all neurons that respond to shades of blue, or to 180° angles. Subsequently, memories are distributed, that is, encoded as bundles of microfeatures (Hebb 1949; Hinton et al. 1986). This neural architecture is an elegant solution to the problem of limited material space of the brain-attic since an individual neuron can contribute to encoding a variety of items. For instance, the same shade-of-red sensitive neurons involved in memorizing red diamonds of a given card set are also involved in memorizing its red hearts, as well as a host of other items that share the same shade of red. Furthermore, the brain is organized so that similar items are distributed across overlapping regions, which facilitates associative recall. By analogy with computer memory types, this feature is often referred to as 'content addressability'. *Content-addressable memory* (CAM) takes data as input, and returns addresses where similar data is stored, while *random access memory* (RAM) takes addresses as input and returns the data stored at those addresses (Hebb 1949; Marr 1969). Grouping similar data improves the performance of CAM, but not that of RAM.

Holmes' BAM reflects an outdated view of the material substrate of memory. At the same time, it is clearly a sort of CAM, and displays a cogent outlook on the functional organization of memory. A BAM-CAM distributed architecture creates a host of potential associations, contingent on one's past experience. For instance, someone who has first encountered a given shade of red through the red diamonds of a particular card set would associate that shade with a particular rhombus shape.

[4] Although not explicitly mentioned in Holmesian canon, 'mind palaces' have become a trope of contemporary versions of Sherlock Holmes. They are featured in the following episodes of the BBC's *Sherlock*: "The Hounds of Baskerville" (S02E2, 2012), "The Empty Hearse" (S03E1, 2014), "The Sign of Three" (S03E2), "His Last Vow" (S03E3), and are the principal plot device in "The Abominable Bride" (Special, 2016). They are also referred to in one episode of CBS's *Elementary*, "The Long Fuse" (S01E8, 2012).

Depending on the context, these associations may just as well draw their attention to features of the environment relevant to their current task or divert it towards irrelevant ones. One can cultivate associations, as recommended by the method of *loci* or Holmes' method. Then, particular visualizations create patterns of co-activation in a BAM-CAM. The relation between the functional-level description (the *locus* an individual memory is kept in, and its relations with other *loci* and memories) and a material-level description (the patterns of co-activation of cliques of neurons) is more subtle than the one-one correspondence suggested by Holmes' BAM description. But Holmes' recommendation to keep "a large assortment [of tools in] perfect order" has nonetheless a possible description at both levels. Furthermore, it offers a first interpretation of the notion of strategic associations.

1.3 Attention and strategic associations

For a "skillful workman" whose trade (like Holmes') is a particular brand of problem-solving, the "most perfect order" is the one that fosters useful associations facilitating said problem-solving. In some cases, expert knowledge is the right tool for that trade. Holmes' extensive readings about past criminal cases have informed him about the kind of evidence found at a variety of crime scenes. Subsequently, he has made a special study of cigar ashes and written a monograph on the topic;[5] and he learned how to identify a newspaper clipping by its paper and the type of its letters. This expert knowledge sensitizes Holmes to features of his environment that he can identify while others cannot. For instance, Holmes spots in *A Study in Scarlet* scattered Trichinopoly ashes on the floor of the abandoned house of Lauriston Garden, and later identifies the murderer upon finding the same ashes in his hotel room. In *The Hound of the Baskervilles*, Holmes peruses words cut up from a newspaper and glued together to compose an anonymous letter, and identifies the shape of their type and the quality of the paper they are printed on, characteristic of the *Times*. Together with other indications, the finding suggests a complete method for identifying the sender.[6]

M. B. and J. Hintikka define the *range of attention* of Inquirer as the set of *yes–no* questions that Inquirer is ready to raise in a given interrogative game (Hintikka and Hintikka 1989). With this parameter, the role of a well-ordered 'brain-attic' in

[5] The monograph titled *Upon the Distinction Between the Ashes of the Various Tobaccos* is mentioned in "A Study in Scarlet", "The Boscombe Valley Mystery" and "The Hound of the Baskervilles".

[6] "We are coming now rather into the region of guesswork," said Dr. Mortimer.
"Say, rather, into the region where we balance probabilities and choose the most likely. It is the scientific use of the imagination, but we have always some material basis on which to start our speculation. Now, you would call it a guess, no doubt, but I am almost certain that this address has been written in a hotel."
"How in the world can you say that?"
"If you examine it carefully you will see that both the pen and the ink have given the writer trouble. The pen has spluttered twice in a single word and has run dry three times in a short address, showing that there was very little ink in the bottle. Now, a private pen or ink-bottle is seldom allowed to be in such a state, and the combination of the two must be quite rare. But you know the hotel ink and the hotel pen, where it is rare to get anything else. Yes, I have very little hesitation in saying that could we examine the waste-paper baskets of the hotels around Charing Cross until we found the remains of the mutilated Times leader we could lay our hands straight upon the person who sent this singular message." (*The Hound of the Baskervilles*, ch. 4).

Holmesian inquiry can readily be interpreted. For instance, in the above examples, expert knowledge is instrumental in improving Holmes' range of attention, allowing him to devise problem-solving strategies. Holmes' method exploits the features of associative memory: associations draw his attention to certain details that in turn trigger other associations and inferences, etc. This creates a feedback loop between attention (and perception) and associative memory that plays a major part in Holmesian deduction. This feedback loop is critical to fully understand what M. B. and J. Hintikka bill as the main task of the "Holmesian logician":

> The crucial part of the task of the Holmesian "logician", we are suggesting, is not so much to carry out logical deductions as *to elicit or to make explicit tacit information*. (Hintikka and Hintikka 1983, pp. 156–157)

Deductive inferences are a rather common interpretation of *eliciting tacit information*, in particular in cognitive psychology and theoretical computer science (cf. Appendix). After M. B. Hintikka's death, the substitution of proof theory for game theory as a theoretical basis for the IMI emphasized this interpretation. But when M. B. and J. Hintikka suggested that eliciting tacit information by questioning could be viewed as a recall procedure, they hinted at another interpretation. In fact, they anticipated on recent findings in cognitive psychology, articulated in the following passage by cognitive psychologist Liane Gabora:

> [A]n associative memory contains information that was never explicitly stored there, but that is implicitly present nonetheless due to the ingenious way one's history of experiences is encoded. [...] [T]his information is accessed in associative thought, and made increasingly explicit in analytic thought, enabling one to go beyond what one knows without resorting to trial and error. (Gabora 2010, p. 5)

Gabora is concerned with providing a scientific explanation of creative thought but analyzes cases that double as instances of problem-solving. Her two main examples are a designer tasked with inventing a new design for a comfortable chair, who solves the problem by conceiving a beanbag chair (Gabora 2010, p. 8); and a penniless ski instructor trying to build a new garden fence at low cost, who solves the problem by recycling old skis as fence posts (Gabora 2010, pp. 9–10). Gabora's explanation of creative thought can be interpreted, relative to 'Holmesian deduction', as the description of a strategic process. First, "analytic thought" encompasses deductive inferences that the interrogative approach readily characterizes strategically. Second, creative thinking also involves associative thought, and we have seen how the architecture of memory supports 'strategic' associations in the context of Holmesian deduction. So understood, Gabora's analysis undermines the argument that there can be no 'logic of discovery' (cf. n. 3) and is therefore relevant to epistemology and philosophy of science. However, there is a stronger game-theoretic sense in which the associations she describes can be part of a strategy. The next section articulates that game-theoretic sense from an example.

2 Visiting our neighbors' brain-attics

2.1 An example: the microwave background of the universe

The Hintikkas' interrogative approach to inquiry is limited to the single-agent case: a 'game against Nature' is in fact a single-agent decision problem under uncertainty, where no strategic interaction takes place between players (Genot 2017). There are however contexts in which strategic interaction occurs if Inquirer addresses her questions not only to Nature but also to informants with their own beliefs and preferences. In some contexts, Inquirer and her informants may have conflicting preferences and informants could prefer not to answer truthfully. We analyzed some strategic inferences for those contexts in Genot and Jacot (2012). In such cases, Inquirer chooses her question so as to *prevent* associations from her informants' part. However, when no conflict between the preferences of Inquirer and those of her informants exists, associations are not only welcome but may be sought after. Crucially, when informants are reasoners, they can form inferential anticipations of their own. Rather than letting Inquirer come up with new questions, they can suggest questions to Inquirer, based on anticipations of what Inquirer could infer from answers to those questions, together with her background knowledge. Let us now consider a real-life example of this, the discovery of the microwave background radiation of the universe by astronomers Arno Penzias and Robert Wilson in 1964, that earned them the Nobel Prize in Physics in 1978.

In the mid-1960s, Russian cosmologists Andrei Doroshkevich and Igor Novikov derived analytically the conclusion that a 'big bang' event would have an observable radiation signature in the microwave spectrum. Their result was published in 1964 in Russian, and soon after in English translation. However, as Penzias later remarked in his Nobel Prize lecture, it "escaped the notice of the [...] workers in [the] field" (Penzias 1978, p. 455). In the same lecture, Penzias gives the following account of the discovery he made with Wilson:

> Shortly [after the Russian publication] P. J. E. Peebles treated the same subject for a different reason. R.H. Dicke had [...] set out to measure the background brightness of the sky at microwave wavelengths. At his suggestion, Peebles began an investigation of the cosmological constraints that might be imposed by the results of such a measurement. Peebles' paper, which was submitted to the *Physical Review* and circulated in preprint form in March of 1965 [...] explicitly delineated the surviving relict radiation as a detectable microwave phenomenon. At about the same time, microwave background radiation was detected at Bell Laboratories and its extragalactic origin established. No combination of the then known sources of radio emission could account for it. Receipt of a copy of Peebles' preprint solved the problem raised [by] this unexplained phenomenon. (Penzias 1978, p. 456)

The actual circumstances of the discovery are much livelier, and of some importance for our discussion. In 1964, Penzias and Wilson were, as part of NASA *Project Echo*, using their radio telescope to detect microwaves emissions from Earth, bounced off satellites *Echo I* and *Echo II*. They recorded an interference pattern, and after ruling out a possible military test as its cause, found that pigeons had been nesting in their antenna. They hypothesized that pigeon droppings could have caused the interference (for reasons unknown), and attempted unsuccessfully to relocate the pigeons. They resolved to shoot them at close range, but the interference pattern persisted, even after the droppings had been cleared. Soon after this incident, Penzias mentioned these issues to a colleague, Bernard Burke, who recalled the topic of Dicke's research and the preprint version of Peebles' paper. Penzias then called Dicke, who was about to build a radio telescope of the type used by Penzias and Wilson, and to measure the microwave signature predicted by Peebles. Dicke recognized the interference pattern described by Penzias as characteristic of this signature and agreed to publish Peebles' analytic results simultaneously with Penzias and Wilson's empirical data.[7]

Penzias, Wilson, Burke, Dicke and Peebles were essentially sharing the same background knowledge. The only relevant difference was that Dicke and Peebles had established analytically that a 'big bang' would have a microwave signature, while Penzias and Wilson were not aware of that consequence. Dicke would have preferred that his own team would validate empirically Peeble's calculations (see n. 7) but his preference was not strong enough to withhold relevant information. Therefore, the discovery is a case of perfect cooperation between inquirers and informants. From the standpoint of the interrogative approach, both teams (Penzias and Wilson, and Dicke and Peebles) could be described as inquirers, as well as informants for one-another. Only Burke has a clear-cut informant role. Even more interestingly, the discovery can be described as the outcome of a strategy profile (a n-tuple of strategies) that includes strategies of essentially the same type, that we will now describe.

2.2 Sagacity and luck

The discovery of the cosmic microwave background is a garden-variety of *serendipity pattern*, a phrase coined by Merton (1948) and referring to "the fairly common experience of observing an unanticipated, anomalous, and strategic *datum* which becomes the occasion for developing a new or extending an existing theory" (p. 506).[8] In recent years, the notion of serendipity has become popular in information studies, but a recent survey of the field by Foster and Ellis (2014) concluded to a lack of consensus about the meaning of the term. On the one hand, information scholars agree that "falling somewhere between accidental and sagacity, serendipity is synonymous with neither one nor the other" [Danzico (2010), p. 16, cited by Foster and Ellis]. On the other hand, there is no consensus about where it stands relative to either. This conceptual situation could be greatly improved by appealing to the Hintikkas' analysis. First, we can interpret Danzico's remark relative to this analysis, as follows: (1) letting 'acciden-

[7] Penzias related the anecdote in an interview led by Ralph Schoenstein (titled "The Big Bang's Echo") for the segment *All Things Considered* aired on National Public Radio (May 17, 2005, accessible online, with transcription, at: http://www.npr.org/templates/story/story.php?storyId=4655517). Penzias reports that upon hanging up the phone, Dicke said to his team: "Guys, we've been scooped".

[8] Merton co-opted the term 'serendipity' from earlier authors and credited Peirce for influencing his definition of serendipity patterns. Barber and Merton (2004) later gave a detailed history of the term before and after Merton's original paper.

tal' refer to the broadening of Inquirer's range of attention in response to unexpected external circumstances, and: (2) letting 'sagacity' refer to her ability to form strategic anticipations based on the inferential role of answers. Second, the interrogative approach readily captures at least one relation between the two, namely circumstances under which chance external events broaden an agent's range of attention, and through associations, result in inferential anticipations that would not otherwise have been possible. In fact, the approach can capture other relations, and in particular strategies that increase the odds that such chance events will occur in contexts where *interaction* with other agents (and their brain-attic) takes place.

Penzias and Wilson's discovery of the cosmic microwave background of the universe offers an example where a serendipitous discovery results from outsourcing one's range of attention. Burke associated Penzias' issue with Peebles' research, and expected that Dicke could suggest relevant questions. And indeed, during the conversation with Dicke, the question *Is the interference characterized by such-and-such pattern?* entered Penzias' range of attention, soon followed by the conditional *If the interference is characterized by such-and-such pattern, then it is caused by a 'big bang' event*, which was a consequence of Peebles' research. A positive answer to the former question yielded (by *modus ponens*) a complete answer to Penzias and Wilson's main research question (*what is the source of the interference pattern recorded by our Horn antenna?*). As we also mentioned, this discovery could also be described from the standpoint of Dicke and his team. In particular, circulating the preprint of Peebles's paper played, for Dicke and Peebles, the same role as talking to Burke for Penzias and Wilson. For both teams, the phone call between Penzias and Dicke resulted in the same *modus ponens* inference to the validation of the 'big bang' model, one team providing the antecedent, and the other the conditional, and the phone call between Penzias and Dicke completing the inference.

In Hintkka's terminology, a conditional such as the above is called a *conclusiveness condition*. More formally, a conditional *If A, then B* is a conclusiveness condition for a question Q if B is one of the potential answers to Q. A simple heuristic principle for answering some question Q is to search one's background knowledge for conclusiveness conditions for Q and ask *yes–no* questions about their antecedent. In some contexts, this heuristic principle can be implemented mechanically (Genot 2017, pp. 16–17). But in the case of the discovery of the cosmic microwave background, Penzias and Wilson followed another heuristic. Namely, they outsourced the process of forming associations between answers to the question they were investigating, and the knowledge available at the time among astronomers, when they talked to Burke. Similarly, they outsourced the inferential process when Penzias talked to Dicke on the phone. This heuristics illustrates why serendipity can be conceived of as the product of both luck and sagacity. As we said, 'sagacity' can be here equated with the ability to form strategic anticipations about the inferential consequences of some 'small' answers, together with some conclusiveness conditions retrieved from one's memory. In the case of Penzias and Wilson, the process of forming such anticipations was partially implemented by someone else (Dicke), and Penzias and Wilson's sagacity was actually to rely on someone else's sagacity.

2.3 Behavioral strategies of inquiry

Cognitive outsourcing of the type just described vindicates the view that serendipitous discoveries fall somewhere between sagacity and luck without reducing to either. In a case like the discovery of the cosmic microwave background, it is possible to describe inquiring agents as almost literally ‘betting’ on the ability of informants to provide additional information—be it explicit or implicit, where the latter includes the products of both memory recalls and deductive inferences (cf. Sect. 1.1). Formally, this betting process is captured by the game-theoretic notion of *behavioral strategy*. A (complete) *pure* strategy for a given player in a given game is a function assigning an action to that player, for every position it is that player’s turn to play. For instance, a strategy for player *White* in the game of chess assigns one move for *White*, to every possible sequence of moves whose last element is a move made by *Black*. A *behavioral* strategy is a sequence of lotteries over moves, equivalent, in our game of chess, to pick a move at random among those that *White* can play after each particular sequence whose last element is a move made by *Black*. *White* can bet symmetrically (or ‘equivocate’) with an equal chance to play any of the moves allowed after *Black*’s; or she may have a particular bias for some moves at some position, and thus a greater chance to play some moves than others at that position. In the limit, the strongest possible bias recommends exactly one move: from a game-theoretic standpoint, pure strategies are degenerate cases of behavioral strategies.

A strategy conducive to serendipitous discovery, such as the one followed by Penzias and Wilson (and to a lesser extent by Dicke and Peebles), is a behavioral strategy with lotteries over informants at given positions of the inquiry. Inferential moves are still deterministic, but informants can provide with unexpected information, and may open unforeseen possibilities for further inferences. Hence, the inquirers’ strategies cannot be represented by complete strategies, as they can become aware of new possibilities for inferential moves when new questions enter their range of attention. Representing the inquirers’ awareness requires resources beyond those of classical game theory. One needs to associate to each position it is one of the players’ turn to play, a representation of the game for that player at that position. And one needs to define the players’ strategy spaces relative to these indexed representations. Extensive games with (un)awareness suitable for representing such situations were first introduced by Halpern and Rego (2006), and we discussed their application to inquiry games (in particular multi-agent cases with uncooperative sources) in Genot and Jacot (2012).

In single-agent (decision-theoretic) reconstruction of information seeking behavior, shifts in agent’s attention must be represented as depending on Nature’s state. If some of these shifts are not deterministic, they must be represented as lotteries over possible states of Nature, that is, chance events. In a multi-agent (game-theoretic) reconstruction, it becomes possible to represent how inquirer can ‘manipulate chance’, that is, deliberately introduce lotteries in her information-gathering behavior. For instance, one could reconstruct Penzias’ conversation with Burke, as part of a behavioral strategy, based on some heuristics. For instance: for every colleague that Penzias would meet, and who would have some extensive knowledge in astrophysics, even outside Penzias and Wilson’s field, randomize between mentioning the interference, or not. Intuitively, the weights of the lottery would represent how orthogonal preferences are weighed against one another (here: the opportunity of gaining insights from informal conversations, vs. the risk of being perceived as work-obsessed and boring).[9] The perspective on serendipity typically adopted by sociologists of science and information scholars is implicitly grounded in decision theory. From this perspective, ‘luck’ is mapped to chance events in the state of Nature, while ‘sagacity’ is mapped to reasoning and decision-making processes. By contrast, a reconstruction explicitly grounded in game theory further constrains ‘luck’, making it part of the strategy selection process, to which is also mapped ‘sagacity’. Then, at least in multi-agent cases, there is a rather precise interpretation of the notion that serendipity holds a middle ground between sagacity and luck. In the next section, we suggest that this game-theoretic interpretation for the multi-agent case is also relevant to the single-agent case.

3 Distributed information

3.1 Distributed knowledge

The information seeking strategies exemplified by Penzias and Wilson, and by Dicke and Peebles, are widespread in academia. Academics have informal talks with knowledgeable colleagues (as the saying goes, ‘pick the brain’ of one another), they present work in progress to seminars and conferences, and they acknowledge feedback from audiences (and sometimes reviewers) when they publish final versions. Knowledge tapped into by those strategies is distributed among us and our colleagues. More generally, knowledge that ϕ is *distributed* in a given group, if ϕ is known by the members of the group after they have pooled together what they know individually. While neglected by mainstream epistemologists, and philosophers and sociologists of science, distributed knowledge has been discussed by formal epistemologists and computer scientists, beginning with Fagin et al. (1995);[10] and perhaps more surprisingly, by sociologists of organizations and management researchers, following the work of Tsoukas (1996, 2003). The issues raised by Tsoukas about business firms generalize to epistemic communities whose members have bounded cognitive resources, as conspicuous in the following passage:

[9] From a decision-theoretic standpoint, the situation would be captured as a one-shot decision problem (or: game vs. Nature), and repeated interaction would thus appear irrelevant to maximizing one’s expected utility. From a game-theoretic standpoint, the maximization of utility must be weighed against the risk of other scientists perceiving the agent as work-obsessed and boring, therefore decreasing over time the chance of interaction (and thus, of getting insights from the scientific community). Repeated interaction and the fact that the interaction between Inquirer and a (strategic) source S_1 may impact the strategy of a source S_2 at a later stage is most naturally captured as a genuine sequential *game*.

[10] Distributed knowledge is related to the better-known notion of *common knowledge*, but is more elusive, because its formal definition is dependent on ontological commitments about possible worlds, whereas common knowledge is not (see Gerbrandy 1998, 1999); and its formal semantic properties require an extension of the standard methods for epistemic logic. Subsequently, some open problems identified early after its introduction (e.g. in Hoek et al. 1999) were only solved recently. Of particular interest are Roelofsen (2007), who presents an appropriate extension of semantic methods; and Ågotnes and Wáng (2015), who characterize the relation of distributed knowledge to common knowledge.

> The organizational problem firms face is the utilization of knowledge which is not, and cannot be, known by a single-agent. Even more importantly, no single agent can fully specify in advance what kind of practical knowledge is going to be relevant. Firms, therefore, are distributed knowledge systems in a strong sense: they are decentered systems, lacking an overseeing 'mind'. The knowledge they need to draw upon is inherently indeterminate and continually emerging; it is not self-contained. Individuals' stock of knowledge consists of (a) role-related normative expectations; (b) dispositions, which have been formed in the course of past socializations; and (c) local knowledge of particular circumstances of time and place. A firm has greater-or-lesser control over normative expectations, but very limited control over the other two. At any point in time, a firm's knowledge is the indeterminate outcome of individuals attempting to manage the inevitable tensions between normative expectations, dispositions, and local contexts. (Tsoukas 1996, p. 11)

Substituting 'firms' with 'academic communities' in the above passage depicts a familiar picture. For instance, relative to the community of astrophysicists circa 1964, expectations that ivy-league Princeton astronomers (Dicke's team) would build a Horn antenna to validate their analytical results are "role-related normative expectations". The habit of talking to colleagues (Penzias and Burke) and of circulating preprints of important papers (Dicke and Peebles) are "dispositions […] formed in the course of past socializations". Finally, the result published by Doroshkevich and Novikov, known to their editors and translators (but lost to Penzias, Wilson, Burke, Dicke and his team), as well as the content of the observations made by Penzias and Wilson and Peebles' analytic results, are all examples of "local knowledge". Notice that Tsoukas' forms of "stock knowledge" overlap: the circulation of Peebles' preprint can be seen as the result of enforcing several "role-related normative expectations" (showcasing Princeton's research leadership and disseminating results for the benefit of the community), instantiating a disposition (as noted), and inducing local knowledge (Burke's, later beneficial to Penzias).

Tsoukas' analysis of distributed knowledge is influenced by F. Hayek's critics of neoclassical economics, in particular the assumption that agents perform economic calculations from full and correct information. Hayek's argument has a parallel in epistemology, due to the assumption that the information sets of agents are closed under logical consequence, also known as the *problem of logical omniscience*.[11] This problem is easily illustrated with, again, the community of astrophysics in 1964. The hypothesis of an expanding universe, together with Einstein's special relativity, entails the existence of a cosmic microwave background. If we apply the semantic concept

[11] For instance, Gerbrandy notes that: "Possible world semantics give rise to a rather distorted notion of belief and knowledge. In particular, […] the belief (and knowledge) of an agent is closed under logical consequence. This does not conform at all to the situation as it actually is: people do not always see the consequences of their beliefs, and know that other people are limited in the same way. […] This mismatch between the semantics and the concept of knowledge and belief is known as 'the problem of logical omniscience'." (Gerbrandy 1999, p. 40) Gerbrandy defends the idealizations of formal semantics with the remark that *information* is closed under logical consequence. J. Hintikka, who offered the first systematic logical analysis of belief and knowledge in the 1960s (Hintikka 1962), also defended the view that it is best understood as a logic of information (Hintikka 2007, Ch. 1–3).

of knowledge, the existence of the cosmic microwave background is thus 'known' by everyone in the community, but also known by everyone to be known by everyone, etc., and is thus not only distributed knowledge, but common knowledge. By contrast, less-than-logically omniscient agents do not expect others to draw the consequences of what they know, and sometimes interact with others in order to obtain assistance figuring out the consequences of what they know. Tsoukas' informal account captures this as, respectively, "normative expectations" and "dispositions [...] formed in the course of past socializations". Therefore, strategies for co-opting the inferential and associative processes of other members of an epistemic community that we described in Sect. 2.3 are consistent with Tsoukas' account of how knowledge is distributed and accessed in an epistemic community of less-than-logically omniscient agents.

3.2 Distributed memory

Since the memory of each individual is also distributed, and "contains information that was never explicitly stored there" (cf. 1.3), there is a continuity between the means by which members of a community tap into knowledge distributed in that community and the means by which individual members of the community access the implicit content of their own memory. So far, we have seen that accessing knowledge distributed in the community involves behavioral strategies (lotteries over informants). We will now argue that behavioral strategies also capture formally the process of accessing implicit information distributed in one's memory.

Given the architecture of human memory, recall is best described as a probabilistic process. Since items in memory are distributed as bundles of microfeatures, there is always a certain probability that recalling item i_1 would trigger recall of some other item i_2 sharing some microfeature with i_1. The probability that i_1 and i_2 are associated during a particular recall episode depends in part on individual history. Hence Holmes' recommendation to keep one's brain-attic in order, lest "knowledge which might be useful [...] is jumbled up with a lot of other things". Methods of memorization, such as the *method of loci*, reinforce associations. They foster potentially useful associations by introducing particular biases in the probabilistic process of recall. More precisely, they alter the odds that a state of Nature at some time t_i is a state of Nature where i_1 and i_2 are recalled together, as the consequence of steps undertaken at some time t_{i-n}. With the notion of behavioral strategy in place, this can be expressed as selecting a particular lottery for a behavioral strategy for recall. Anticipating on that conclusion, we offered it as a first interpretation of the notion of strategic use of memory.

So far, we have only considered probabilistic weights that result from pro-active (*ex ante*) decisions and choices, made prior to encountering particular problem-solving situations, and based on anticipations of what such situations might bring. However, the architecture of memory also supports reactive (*in medias res* or *post hoc*) associations in situations that one has not fully anticipated. Let us first recall Gabora's examples, which are both cases of creative thinking and problem-solving, and consider how Gabora characterizes the neural correlate of creative insight in such situations:

> [A]nalytic thought requires a state of focused attention, and associative thought requires a state of defocused attention. Creativity involves not just the capac-

> ity for both but the capacity to spontaneously shift between them according to the situation, referred to as contextual focus. […] [C]ontextual focus is explicable at the level of cell assemblies […] composed of neural cliques, some of which respond to situation-specific aspects of an experience, and others of which respond to general or abstract aspects. […] Those neural cliques that respond to atypical features of the situation, and thus that are activated in associative but not analytic thought, are referred to as *neurds*. […] Leakage of information from outside the problem domain occurs in a manner that is a priori unpredictable because it involves unearthing associations that exist due to the overlap of items in memory that may not have been previously noticed. Because memory is distributed, coarse-coded, and content-addressable, items encoded previously to neurds are superficially different from the present situation yet share aspects of its deep structure. Therefore, the recruitment of neurds may foster associations that are seemingly irrelevant yet potentially vital to the creative idea. (Gabora 2010, p. 13)

In Gabora's example of the beanbag chair (mentioned in Sect. 1.3), the typical features of a chair include 'having a back' and 'having legs' and the atypical features '[adapting/not adapting] to one's body'. In a state of focused attention, there is a high probability of recalling examples of other objects sharing some of the first features, like benches and armchairs. In a state of defocused attention, associations with, and recall of, objects that share atypical features of a chair, such as a beanbag, become more probable. The same holds, *mutatis mutandis*, for fence posts and decommissioned skis. Contextual focus is for the most part under the control of the subject, who can let her attention drift, or on the contrary re-focus it. Since alterations of contextual focus affect recall in real-time, they are tantamount to selecting particular behavioral strategies for recall.[12] This description makes conspicuous that, from a strategic standpoint, defocusing one's attention and randomizing one's sources of information in an epistemic community belong to a continuum of strategies for accessing distributed knowledge.

Concluding remarks

In this paper, we have argued that means for accessing information distributed in the brain (associative memory) and in a community (distributed knowledge) belong to a continuum of strategies. In the light of hindsight, this should be no surprise. Knowledge distributed within a community is also distributed across the associative memories of the individual members of the community. Hence, strategies for accessing distributed knowledge should be expected to scale up strategies for accessing the content of one's individual memory. We illustrated this point with the example of a strategy leading to a serendipitous discovery, that scaled up strategies usually conducive of creative solutions in the single-agent case. The conceptual payoff of homing in on this example

[12] More precisely, both contextual focus and defocused attention induce different lotteries for the activation of cell assemblies, with focused (defocused) attention lowering (increasing) the probability that cliques of neurds are activated.

was the precise characterization of the relevant notion of ‘strategy’. Specifically, we showed that it is captured by the notion of *behavioral strategy* at the multi-agent level. We then came full circle, and justified our early identification of the method of *loci* as a strategy, by arguing that it supports real-time behavioral strategies for recall in the single-agent case.

We have built our argument on a cost-benefit analysis of distributed knowledge rather than on a semantic one, on the grounds that the former is sensitive to the resource-boundedness of real-life agents. Cost-benefit analyses capture the down-to-earth reality of laboratory life. For instance, in order to rediscover Doroshkevich and Novikov’s analytic result, and to build a radiotelescope, Dicke’s team had to face salary and infrastructure costs. These expenses could have been saved with a better management of the knowledge distributed in the astrophysics community of the time. Again, our suggestion that these considerations are relevant to epistemology is not a novelty, and is similar to theories of inquiry proposed by American pragmatists (Peirce, Dewey, or more recently Levi). This thread could be further pursued into a refined analysis of distributed knowledge, considering that on the one hand pragmatic theories of inquiry and on the other hand sociology of organizations and management research overlap in both subject-matter and methodology.

The Hintikkas’ approach was an analytic (or conceptual) model of inquiry, scientific or otherwise, whose ambition was to make discovery amenable to logical and strategic inquiry. We have vindicated their approach and argued that the non-analytic component of discovery, that is, analogical processes, is also amenable to the same analysis. The Hintikkas’ approach could also be a starting point to computational models of discovery. We have suggested (in Genot 2017) that discovery algorithms based on pattern recognition on data streams first proposed by Simon (1973) could capture some of the dynamics of an agent’s range of attention. The picture we have given of this dynamics in Sect. 1 of the present article is more sophisticated. Since it supervenes on memory, each input in a data stream can elicit recollections which may in turn suggest new patterns.

Another promising thread of investigation would be the implementation of our theoretical model into a *multi-agent* computational model. In this model, an agent would be characterized by a memory representing the agent’s background knowledge and a range of attention comprising the presuppositions of questions she is ready to ask. The memory would be represented so as to capture probabilistic relations between memory items. At each step of a discovery process at which a question is selected by an agent, she can also select a source: her memory, the environment or other agents. Obtaining answers from any of those sources allows her to draw further conclusions. Obtaining answers from other agents can have such effects as broadening her range of attention.

It is however unclear whether a multi-agent computational model would improve upon the conceptual analysis of multi-agent inquiry captured by the Hintikkas’ analytic model. For instance, the analytic model is not predictive. Hence, a computational model could not serve as a proxy for empirical validation of the model’s predictions through simulations. However, the model could be used to explore ways to generate ‘serendipitous’ recommendations in a network of agents. A possible application would be computer-assisted collaborative inquiry learning, where artificial agents could assist human learners by making recommendations based on anticipations of the learners' strategies. The practical implementation of this application could draw inspiration from algorithms developed for recommender systems that make unexpected but useful recommendations based on the user's preferences. These algorithms have been recently shown to outperform other methods including those that introduce random variations (see Adamopoulos and Tuzhilin 2014, 2015).

The last remark suggests some possible relation with the recent epistemological trends that we mentioned at the end of our introduction. The notion of distributed knowledge that we have investigated has explicit ties with the formal (semantic) notion of distributed and common knowledge investigated by logicians and computer scientists in the 1980s. We argued that, *modulo* the Hintikkas' model, the informal notion of distributed knowledge introduced by sociologists of organizations can solve some of the difficulties associated with the formal notion (in particular the vexing problem of logical omniscience). What remains to be determined is the relation with a third and more recent understanding of 'distributed' knowledge as e.g., characterized by K. Michaelian and J. Sutton, namely "distributed across heterogeneous systems combining neural, bodily, social, and technological resources." (Michaelian and Sutton 2013, p. 1). Our argument has covered in part the 'neural' and 'social' aspects of distribution but did not consider either the 'bodily' or 'technological' ones. We must admit that we have little to say about the former. However, the relation we just alluded to between algorithms for recommender systems and strategies favoring serendipitous discoveries may foreshadow a possible extension of our argument to the technological dimension.

Acknowledgements E. J. Genot's position was partially funded by Vetenskapsrådet, within the project *Knowledge in a Digital World* (Principal investigator Erik J. Olsson).

Appendix: Deductive versus non-deductive problems

A deductive problem. A canonical example of a deductive problem, initially proposed by Levesque (1986), is the following:

> Imagine a world where:
>
> *Jack is looking at Ann but Ann is looking at George.*
>
> and
>
> *Jack is married but George is not.*
>
> I claim that in such a world, we have that:
>
> *A married person is looking at an unmarried one.*
>
> To see why, think of Ann's situation: if she is married, then because she is looking at George, she is the married person looking at an unmarried one; on the other

> hand, if she is not married, because Jack is looking at her, he is the married person looking at an unmarried one. Either way, the conclusion follows, so the information is already implicit in the first two sentences. (Levesque 1986, p. 85)

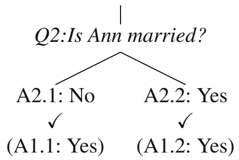
Levesque proposed his problem as a special case of a general problem for artificial intelligence (how to program computers to process implicit information and draw explicit conclusions). It was later co-opted by psychologists Toplak and Stanovich (2002) under the following form:

> *The Married Problem.* The married problem [is] borrowed from Levesque […] and reads as follows:
>
> Jack is looking at Ann but Ann is looking at George. Jack is married but George is not. Is a married person looking at an unmarried person?
> A) Yes B) No C) Cannot be determined
>
> What Levesque defined as implicit information in this problem is analogous to what we, following Shafir (1994), have earlier discussed as fully disjunctive processing of problem components. That is, one must consider the disjuncts of Ann's marital status, which is the implicit information, in order to derive the correct conclusion. If Ann is married, then the answer is "Yes" because she would be looking at George who is unmarried. If Ann is not married, then the answer is still "Yes" because Jack, who is married, would be looking at Ann. The correct solution to this problem can be derived only by using a disjunctive strategy. (Toplak and Stanovich 2002, p. 201)

Shafir (1994), the inspiration for the controlled experiment carried by Toplak and Stanovich, proposed that "thinking through disjunctions" amounts to exploring the branches of a decision tree. Toplak and Stanovich further decomposed the process in the *Married Problem* as: (1) raising an instrumental question (*Is Ann married?*); (2) no answer being available, reasoning by cases; and: (3) drawing the same conclusion in the two cases that need be considered. The process is represented in the decision tree below, assuming that the domain of discourse is a subset of the domain of persons:

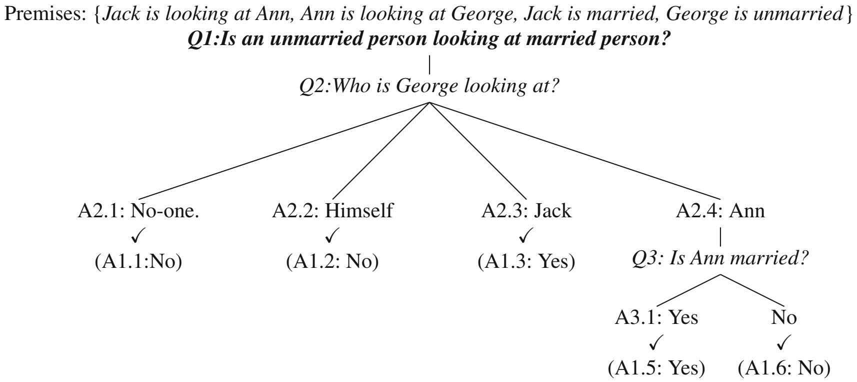
Premises: {*Jack is looking at Ann, Ann is looking at George, Jack is married, George is unmarried*}

Q1: Is a married person looking at an unmarried one?



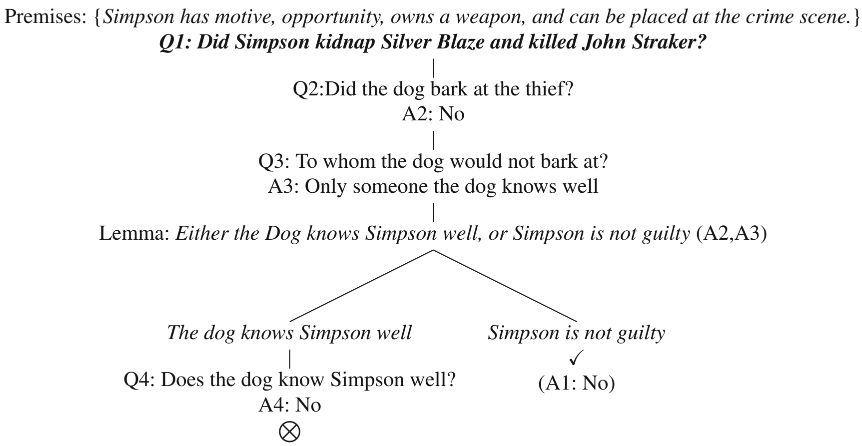
The "deductive strategy" is possible here because, whatever the answer A2.x to Q2 is, the answer A1.x to Q1 is the same. There is no need to actually obtain additional information.

An easy non-deductive problem. A non-deductive variant of the married problem is easily obtained, assuming the same premises, but substituting the principal question with: *Is an unmarried person looking at a married person?* This variant cannot be solved without actually obtaining answers to instrumental questions. However, there is an optimal interrogative strategy to solve it, displayed below, under the slightly stronger assumption that the domain of discourse is the subset of the domain of persons that contains only Jack, Ann and George:



The strategy is optimal, because it does not ask more questions than necessary, and is better than the strategy selecting $Q3$ before $Q2$. Indeed, asking $Q2$ warrants an answer to $Q1$ in three out of four cases, while $Q3$ does so only conditionally upon receiving the answer "Ann" to $Q2$, and provides redundant information in the other cases. The problem is 'easy' insofar as $Q2$ and $Q3$ are expressed with vocabulary occurring in the premises in $Q1$ alone, and could thus be generated mechanically from them.

A harder non-deductive problem. A harder non-deductive problem is solved by Sherlock Holmes in *Silver Blaze*, in the process of identifying the thief of a race horse (Silver Blaze) and the killer of the horse's trainer (John Straker). The police hold a suspect (Fitzroy Simpson) that circumstantial evidence places at the crime scene. The police believe that Straker expected a nightly visit from Simpson, acted upon the suspicion, and met his end at Simpson's hand. The night Silver Blaze disappeared, the horse was kept in a stable together with a watchdog. Holmes' reasoning is represented by the following interrogative-deductive strategy:



Questions $Q2$ and $Q3$ are motivated by Holmes' background knowledge (*the only person a trained watchdog does not barked at is someone the dog knows well*). Holmes obtains $A2$ from Inspector Gregory, in charge of the case; and $A3$ and $A4$, from his own memory. Specifically, $A3$ is part of Holmes' background knowledge, which makes $Q3$ somewhat of a rhetorical question. And $A4$ is obtained by recalling a newspaper account of a visit that Simpson made the evening prior to the crime (a stable boy had let the dog at Simpson, and the dog had indeed barked at him). Inspector Gregory, who does not recall that fact, fails to include it in the premises he considers, and subsequently to ask $Q2$–4. The difficulty of the problem comes from the fact that $Q2$–4 cannot be obtained mechanically from the premises, which do not feature the appropriate vocabulary. $Q2$–4 can only be asked based on associations between the premises and some other content stored in memory.

References

Adamopoulos, P. & Tuzhilin, A. (2014). On over-specialization and concentration bias of recommendations: Probabilistic neighborhood selection in collaborative filtering systems. In *Proceedings of the 8th ACM conference on recommender systems* (pp. 153–160). ACM.

Adamopoulos, P., & Tuzhilin, A. (2015). On unexpectedness in recommender systems: Or how to better expect the unexpected. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *5*(4), 54.

Ågotnes, T. & Wáng, Y. N. (2015). Resolving distributed knowledge. In R. Ramanujam (Ed.), *Proceedings of 15th conference on theoretical aspects of rationality and knowledge, TARK 2015, Carnegie Mellon University, Pittsburgh, USA, June 4–6, 2015* (pp. 31–50). EPTCS.

Barber, E., & Merton, R. K. (2004). *The travels and adventures of serendipity*. Princeton: Princeton University Press.

Beth, E. (1955). Semantic entailment and formal derivability. *Mededelingen der Koninklijke Nederlandse Akademie van Wetenschappen*, *18*, 309–342.

Churchland, P . S., & Sejnowski, T . J. (1992). *The computational brain*. Cambridge, MA: MIT Press.

Danzico, L. (2010). The design of serendipity is not by chance. *Interactions*, *17*(5), 16–18.

Fagin, R., Halpern, J . Y., Moses, Y., & Vardi, M. (1995). *Reasoning about knowledge*. Cambridge, MA: MIT Press.

Foster, A. E., & Ellis, D. (2014). Serendipity and its study. *Journal of Documentation*, *70*(6), 1015–1038.

Gabora, L. (2010). Revenge of the "neurds": Characterizing creative thought in terms of the structure and dynamics of memory. *Creativity Research Journal*, *22*(1), 1–13.

Genot, E. J. (2017). Strategies of inquiry. *Synthese*. https://doi.org/10.1007/s11229-017-1319-x.

Genot, E. J., & Jacot, J. (2012). How can yes-or-no questions be informative before they are answered? *Episteme*, *9*(2), 189–204.

Genot, E. J., & Jacot, J. (2017). Logical dialogues with explicit preference profiles and strategy selection. *Journal of Logic, Language and Information*. https://doi.org/10.1007/s10849-017-9252-4.

Gerbrandy, J. (1998). Distributed knowledge. In I. J. Hulstijn & A. Nijholt (Eds.), *Twendial98: Formal semantics and pragmatics of dialogue* (pp. 111–124). Enschede: Universiteit Twente.

Gerbrandy, J. D. (1999). *Bisimulations on planet Kripke*. ILLC Dissertation Series, Amsterdam.

Halpern, J. Y., & Rego, L. C. (2006). Extensive games with possibly unaware players. In *AAMAS '06: Proceedings of the 5th international joint conference on autonomous agents and multiagent systems, Hakodate, Japan* (pp. 744–751).

Hebb, D. O. (2002). *The organization of behavior: A neuropsychological theory*. Hove: Psychology Press. (First ed. 1949).

Hintikka, J. (1962). *Knowledge and belief*. Ithaca (NY): Cornell University Press.

Hintikka, J. (2007). *Socratic epistemology*. Cambridge: Cambridge University Press.

Hintikka, J., Halonen, I., & Mutanen, A. (1999). Interrogative logic as a general theory of reasoning. In *Inquiry as inquiry: A logic of scientific discovery* (pp. 47–90). Dordrecht: Kluwer.

Hintikka, J., & Hintikka, M. B. (1983). Sherlock Holmes confronts modern logic: Toward a theory of information-seeking through questioning. In U. Eco & T. A. Sebeok (Eds.), *The sign of three: Peirce, Dupin, Holmes*. Bloomington: Indiana University Press.

Hintikka, J., & Hintikka, M. B. (1989). Reasoning about knowledge in philosophy: The paradigm of epistemic logic. In *The logic of epistemology and the epistemology of logic* (pp. 17–35). Dordrecht: Kluwer.

Hinton, G., McClelland, J., & Rumelhart, D. (1986). Distributed representations. In D. Rumelhart, J. McClelland & The PDP Research Group (Eds.), *Parallel distributed processing* (Vol. 1, pp. 77–109). Cambridge, MA: MIT Press.

Levesque, H. J. (1986). Making believers out of computers. *Artificial Intelligence*, *30*(1), 81–108.

Marr, D. (1969). A theory of cerebellar cortex. *Journal of Physiology*, *202*, 437–470.

Merton, R. K. (1948). The bearing of empirical research upon the development of social theory. *American Sociological Review*, *13*(5), 505–515.

Michaelian, K., & Sutton, J. (2013). Distributed cognition and memory research: History and current directions. *Review of Philosophy and Psychology*, *4*(1), 1–24. https://doi.org/10.1007/s13164-013-0131-x.

Penzias, A. A. (1978). The origin of elements. In S. Lundqvist (Ed.), *Nobel lectures, physics 1971–1980* (pp. 444–457). Singapore: World Scientific Publishing Co.

Philosophical Issues. (2014). *Special issue on extended cognition* (Vol. 24(1)). London: Wiley.

Popper, K. R. (1992). *The logic of scientific discovery*. London: Routledge. [First ed. 1935 (in German)].

Reichenbach, H. (1938). On probability and induction. *Philosophy of Science*, *5*(1), 21–45.

Roelofsen, F. (2007). Distributed knowledge. *Journal of Applied Non-Classical Logics*, *17*(2), 255–273.

Shafir, E. (1994). Uncertainty and the difficulty of thinking through disjunctions. *Cognition*, *50*(1), 403–430.

Simon, H. A. (1973). Does scientific discovery have a logic? *Philosophy of Science*, *40*(4), 471–480.

Toplak, M. E., & Stanovich, K. E. (2002). The domain specificity and generality of disjunctive reasoning: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology*, *94*(1), 197.

Tsoukas, H. (1996). The firm as a distributed knowledge system: A constructionist approach. *Strategic Management Journal*, *17*(S2), 11–25.

Tsoukas, H. (2003). Do we really understand tacit knowledge? In M. Easterby-Smith & M. Lyles (Eds.), *Handbook of organizational learning and knowledge* (pp. 410–427). Oxford: Blackwell.

van der Hoek, W., van Linder, B., & Meyer, J.-J. (1999). Group knowledge is not always distributed (neither is it always implicit). *Mathematical Social Sciences*, *38*(2), 215–240.

Yates, F. A. (1966). *The art of memory*. London: Routledge and Kegan Paul.