On relative ignorance

Valentin Goranko

Abstract

I discuss relative ignorance of an agent with respect to the knowledge or ignorance of other agents. It turns out, not surprisingly, that even the two-agent case is quite complex and generates a rich variety of naturally arising non-equivalent operators of relative ignorance. In this paper I explore these in a more systematic way and put together several simple, though technically laborious, observations about their inter-relations. For the technical proofs of these I employ the software tool MOLTAP, which implements, inter alia, tableaux for the underlying multi-agent epistemic logic.

"Against logic there is no armor like ignorance"

(Laurence J. Peter)

1 Introduction

1.1 Knowledge and ignorance

In the context of epistemic reasoning and logics, the word ignorance is used as a term indicating lack of knowledge, with several variations. An agent is said to be ignorant of p, if the agent does not know that p holds (regardless of whether it does), whereas the agent is ignorant of the fact that p if p holds but the agent is ignorant of that; and the agent is ignorant whether p, if the agent is ignorant both of p and of ¬p, etc.

Thus, formally ignorance can be represented by various epistemic operators definable in terms of the knowledge operator K. For instance, the ‘ignorance whether’ operator I is formalised as Ip = ¬Kp ∧ ¬Kn¬p.

The logical formalisation of ignorance has been an attractive topic for a long time and a number of logicians and philosophers have studied and re-discovered its logical properties, often in ignorance of previous works on the topic.

With few exceptions, mentioned further in the references on related work, most logical studies of ignorance so far have focused on the single-agent case. In that case, when the underlying knowledge operator is S5, i.e. satisfies truthfulness and both

1On the other hand, traditionally, knowledge is modelled in epistemic logic as lack of uncertainty – the agent knows something if it is true in all worlds (states of affairs) that the agent considers epistemically possible. Yet, ignorance and uncertainty are different concepts and the former has a certain negative connotation in the everyday language. So, I find its use somewhat misleading, but will nevertheless follow the established tradition.

2Here I am borrowing Kit Fine’s pun phrase from a footnote remark in [11], which, to some extent, was also self-referential.
positive and negative introspection, the range of derivable ignorance operators is small and simple to describe. When negative introspection is not assumed, things are more complicated, as observed, inter alia, by Fine in [11], which studies the hierarchy of iterated ignorance operators, i.e. ignorance, ignorance of ignorance, etc. In particular, Fine shows that if knowledge is assumed to be an S4 modality, then that hierarchy stabilises on level two, because it is possible to be ignorant of one’s own ignorance, but “it is impossible to know of one’s second order ignorance”, so second-order ignorance is inevitably third-order ignorance, etc. For weaker knowledge operators, however – for instance, in the case of $T$, when only truthfulness of knowledge is assumed – Fine shows that it may extend infinitely (just like the hierarchy of iterated knowledge) without stabilisation up to logical equivalence.

As one can expect, the case of multi-agent ignorance is considerably more complicated. In this paper (rather, a longer technical note) I look at relative ignorance of agents with respect to the knowledge of other agents. It turns out, not surprisingly, that even the two-agent case is complex enough to go quickly out of hand and warrants a more-detailed and systematic study, as the technical results presented here and a number of so far open technical problems mentioned at the end indicate. In particular, even in the simplest case, of S5 knowledge modalities for each agent, a rich variety of meaningful relative ignorance operators arises. In particular, I identify here a dozen of pairwise non-equivalent formal notions of relative ignorance of one agent with respect to another and show that these notions have generally different logical behaviour with respect to their iterations: some stabilise up to equivalence after 2, 3, or at most 4 applications, whereas others generate infinite families of apparently pairwise logically incomparable iterations. One of the latter cases is the 2-agent operator expressing agent a’s ignorance about agent b’s ignorance about a proposition $p$. This observation is in sharp (but not unexpected) contrast with the above-mentioned collapse of the hierarchy of single-agent’s higher-order ignorance based on S4 knowledge, noted in [11].

The goal and contributions of the present paper are fairly modest: it aims to identify and offer a more systematic look at the various shades of relative inter-agent ignorance and to put together several simple, though technically laborious, observations about their inter-relations. The web-based version of the Modal Logic Tableaux tool MOLTAP [35] proved to be a very useful technical tool for that purpose, saving hours of routine but arduous formal semantic or deductive reasoning (which I would not have done by hand, anyway).

Contingency, ignorance and knowing whether: brief historical and bibliographic notes

The ignorance operator $I$ has been noted and studied extensively for a long time in a purely modal setting, as the contingency modality: $Ip = \Box p \land \Box \neg p$. Formal study of modal operators and logics of contingency and its negation (non-contingency, analyticity) goes back to work by Montgomery and Routley in the late 1960s, [22], [23], [24], followed by Mortensen [25], Cresswell [5], Kuhn [21], Humberstone [17], [18], [19], Demri [6], Zolin [36], Pizzi [27], [28], and others. In particular, the contingency operator has been studied and axiomatized on its own for various natural classes of Kripke frames (i.e., for K, T, K4, S4, S5, etc.) in some of these works.
In epistemic setting, the contingency operator, interpreted as ignorance has not been as prominent as its dual, non-contingency operator, epistemically meaning knowing whether. Both are mentioned explicitly, but not much explored, already in Hintikka’s seminal book [16]. Later, van der Hoek and Lomuscio [32], [33] study and axiomatize the ignorance operator for its own interest and importance, whereas Steinsvold in [30] relates the logic of ignorance LB with the logic of topological border\(^3\). See [9] for a comprehensive work and discussion relating contingency, knowing whether and ignorance, the already mentioned recent paper [11] which explores the formal properties of higher-order (iterated) ignorance, as well as the recent [15] where beliefs-based variations of knowing whether are discussed in a multi-agent setting.

There are also some studies of ignorance in theory of information processing and in decision theory (esp. in theories of rational choice, e.g. [4], [26], [2]) and a few more philosophically motivated related works, incl. some on ‘pluralistic ignorance’ [3], [29], as well as the recent book [7]. For a recent discussion on the topic and some of the above cited works, see [10].

All works mentioned above explore almost exclusively single-agent ignorance. Most related formal studies of multi-agent knowledge and ignorance come either from Game theory or from AI (e.g., [1], [20], [12]) and Computer Science (esp., distributed computing and multi-agent systems), including [14], [13], [31], but only treating ignorance in the context of multi-agent ‘only knowing’.

2 Preliminaries

2.1 Single agent knowledge and ignorance

The single-agent knowledge modality will be denoted, as usual, by \(K\varphi\), meaning ‘the agent knows that \(\varphi\) (is true)’. I will denote its dual by \(\hat{K}\varphi\), with intuitive reading ‘(the truth of) \(\varphi\) is consistent with the agent’s knowledge’. The standard language of single-agent propositional epistemic logic EL is defined as usual, as well as its possible worlds semantics, incl.: epistemic models, truth at a possible world in an epistemic model, satisfiability, validity, logical consequence and equivalence of formulae in EL. In particular: \(\models \varphi\) denotes the claim that \(\varphi\) is valid; \(\models \varphi \rightarrow \psi\) means that \(\psi\) follows logically from \(\varphi\), equivalently that \(\models \varphi \rightarrow \psi\); and \(\varphi \equiv \psi\) means that \(\varphi\) and \(\psi\) are logically equivalent, i.e. that \(\varphi \models \psi\) and \(\psi \models \varphi\). For further details on the basics of EL, see e.g. [8] or [34].

Following Fine’s taxonomy in [11], I will consider several versions of agent \(a\)’s individual ignorance. However, I will afford some slight deviations from Fine’s terminology, as follows.

- \(\neg K_a p\): a is ignorant that \(p\).

There are two cases to be considered here, depending on whether \(p\) holds or not:

**Pseudo-ignorance:** \(\neg p \land \neg K_a p\). Assuming that knowledge is truthful, this essentially says ‘\(\neg K_a p\) holds because \(\neg p\) holds’. In this case the description

\(^3\)A similar observation and a discussion regarding the notion of borderline is made in [11] in apparent ignorance of [30].
‘a is ignorant that p’ is somewhat misleading, as it is unreasonable, or at least unfair, to declare an agent ignorant of something that is false and therefore cannot be truthfully known. So, when \( \neg p \land \neg K_a p \) holds I will say (for lack of better term) that ‘a is pseudo-ignorant of p’. Again, assuming that knowledge is truthful, this is equivalent to \( \neg p \), so this case will not be of further interest for us. Likewise for the logically stronger case when \( K_a \neg p \land \neg K_a \neg p \), which is equivalent to \( K_a \neg \neg p \).

**True ignorance:** \( p \land \neg K_a p \): a is ignorant of the fact that p. This is the case which is of main interest for us.

Still, considering \( \neg K_a p \) alone makes sense when the truth of p is unknown or not specified. In this case one can say that ‘a considers it epistemically possible that \( \neg p \)’. Respectively, when \( \neg K_a \neg p \) holds, one can say that ‘a considers it epistemically possible that p’.

- a is (first-order) ignorant whether p: \( l_a p := \neg K_a p \land \neg K_a \neg p \).
- This is the most common notion of ignorance, at least amongst those studied in formal logical setting. Note that, assuming that \( K_a \) is truthful, i.e. satisfying the axiom T, \( l_a p \) is equivalent to \( (p \land \neg K_a p) \lor (\neg p \land \neg K_a \neg p) \), i.e. is an exclusive disjunction of the two possible true ignorance cases.

I will be denoting the dual of \( l_a \) by \( W_a \), for ‘knowing Whether’, i.e.

\[ W_a p := \neg l_a \neg p \equiv K_a p \lor K_a \neg p. \]

Besides the ignorance operators mentioned above, Fine [11] considers a variety of ‘higher-order ignorance’ operators, including:

- \( l_a^1 p := l_a p \land \neg K_a l_a p \): ‘a is Rumsfeld-ignorant of p’.
- \( l_a^2 p := l_a l_a p \): ‘a is second-order ignorant whether p’.
- \( l_a^3 p := l_a l_a l_a p \): ‘a is third-order ignorant whether p’, etc.

As mentioned earlier, when \( K_a \) is an S4 operator (or stronger), as shown in [11], the three cases listed above turn out to be equivalent. Moreover, assuming \( K_a \) to be an S5 operator, more equivalences hold that simplify matters further, e.g. \( l_a p \equiv K_a l_a p \).

### 2.2 Basics of the multi-agent epistemic logic MAEL

The multi-agent epistemic logic MAEL\(^n\) involves a set of n agents \( A = \{a_1, \ldots, a_n\} \) and knowledge operators \( K_{a_1}, \ldots, K_{a_n} \) associated with each of them. Multi-agent epistemic models for MAEL\(^4\) are Kripke models of the type \( \mathcal{M} = \langle W, \sim_1, \ldots, \sim_n, V \rangle \), where \( W \) is a non-empty set of ‘epistemically possible worlds’, \( V \) is a valuation function over a fixed set of atomic propositions \( AP \), and for each \( i \), \( \sim_i \) is the epistemic indistinguishability relation of the agent \( a_i \), here assumed to be an equivalent relation. Then the semantics of \( K_{a_i} \) is given as usual: \( \mathcal{M}, w \models K_{a_i} \varphi \) iff \( \mathcal{M}, u \models \varphi \) for every \( u \in W \) such that \( w \sim_i u \).

\(^4\)I am tempted to use ‘a is agnostic about p’ for \( \neg K_a p \), but I will resist that, to avoid the misleading association with the theological meaning of agnosticism as not just lack, but impossibility, of knowledge.”
2.3 The Muddy Children scenario

I will use the well-known Muddy Children scenario (cf. e.g. [8]) for the case of 3 children, hereafter denoted MC\(^3\), to illustrate the various operators of relative multi-agent ignorance\(^5\) and to generate counter-examples of some non-valid consequences between them.

The case of 3 Muddy Children, further denoted by MC\(^3\). Three children, A, B, C, are playing in the muddy backyard and some of them have soiled their faces with mud. Each of them can see the other’s faces but not their own. The father comes, looks at them, and says “Some of you have muddy faces. If any of you knows whether your face is muddy, you must step forward and confess that”, etc. the story goes. I will use \(M_X\) to denote the proposition stating that \(X\) is muddy (i.e., has a muddy face), for \(X \in \{A, B, C\}\).

3 On relative multi-agent ignorance

I assume here that each agent’s knowledge is truthful, positively introspective, and negatively introspective, i.e. satisfies the S5 axioms.

As it is well known (cf. e.g. [8] or [34]), the hierarchy of iterated mutual knowledge, even in the case of just 2 agents, \(a\) and \(b\), viz. \(K_a p, K_a K_b p, K_a K_b K_a p,\) etc. is generally strict and infinite up to logical equivalence in S5\(^2\) and leads in the limit to the notion of ‘common knowledge’. Much less is known from the literature about relative ignorance, so let us look at that notion more systematically.

By analogy with multi-agent group knowledge, one can define versions of group ignorance, for any group of agents, as the conjunction of the respective versions of individual ignorance. In particular, if \(A\) is a set of agents, then the first-order group ignorance in \(A\) whether \(p\) is defined as \(I_A p := \bigwedge_{a \in A} I_a p\), whereas the group Rumsfeld-ignorance in \(A\) of \(p\) is defined as \(I^R_A p := \bigwedge_{a \in A} I^R_a p\). These, however, are hardly more interesting and illuminating than their single-agent versions, so I will not consider them hereafter\(^6\).

More interestingly, a rich variety of natural notions of relative knowledge and ignorance between agents arises, of which I will only consider here the cases involving just 2 agents, \(a\) and \(b\). To each notion of relative knowledge there also corresponds a notion of relative ignorance, obtained by taking negation, and vice versa – every notion of relative ignorance generates likewise a corresponding notion of relative knowledge. Here I will be mainly interested in the variety of notions of relative ignorance.

To begin with, let us distinguish again pseudo-ignorance \(\neg p \land \neg K_a p\) from true ignorance \(p \land \neg K_a p\), with non-knowledge \(\neg K_a p\) as a unifying concept. Applying each of the latter two, as well as the operator \(I_a\) of ‘ignorance whether’, to the knowledge or ignorance of another agent generates an unfeasibly large variety of combinations. I

\(^5\)It is somewhat notable that all these operators can be illustrated and distinguished in such a simple epistemic scenario.

\(^6\)On the other hand, meaningful notions of multi-agent ignorance analogous to distributed and common knowledge are potentially quite interesting, but also quite challenging and I leave them for further work.
consider here a representative selection of a dozen of the most natural, in my view, versions of 2-agent relative ignorance operators, i.e. an agent’s ignorance about another agent’s knowledge or ignorance, where I will be using the notation $O_1/O_2$ to denote a composite operator defined by $O_1/O_2p := O_2p \land O_1O_2p$.

1. ($\neg K_aK_bp$) ‘a is ignorant of b’s knowledge of p’: $\neg K_aK_bp$.
   A natural strengthening is $p \land \neg K_aK_bp$.
   For example, in the 3 Muddy Children scenario $MC^3$, $\neg K_AK_BM_A$ holds.
   Moreover, if A is muddy then $M_A \land \neg K_AK_BM_A$ holds, too.

2. ($\neg K_a\neg K_b$) ‘a is ignorant of b’s ignorance of p’: $\neg K_a\neg K_bp$.
   For example, in $MC^2$, $\neg K_A\neg K_BM_A$ holds.

3. ($\neg K_a/\neg K_b$) ‘a is ignorant of the fact that b knows p’:
   $\neg K_a/\neg K_bp := \neg K_bp \land \neg K_bK_bp$.
   For example, if A is muddy then $\neg K_A/\neg K_BM_A$ holds. Likewise, if only B is muddy, after the father’s announcement $\neg K_A/\neg K_BM_B$ holds, too.

4. ($\neg K_a/\neg K_b$) ‘a is ignorant of the fact that b is ignorant of p’:
   $\neg K_a/\neg K_bp := \neg K_bp \land \neg K_a\neg K_bp$.
   For example, in $MC^3$, if A is not muddy then $\neg K_A/\neg K_BM_A$ holds. Likewise, if A and B are muddy and C is clean, then, after the father’s announcement $\neg K_A/\neg K_BM_B$ holds.

5. ($\neg K_aW_b$) ‘a is ignorant of b’s ignorance whether p’: $\neg K_aW_bp$.
   also equivalent to $K_aW_b$.
   For example, in $MC^3$, $\neg K_AW_BM_A$ holds, because $K_AW_BM_A$ holds. Also, if B is muddy and C is clean, $\neg K_AW_BM_B$ holds.

6. ($\neg K_aW_b$) ‘a is ignorant of b knowing whether p’:
   $\neg K_aW_bp \equiv \neg K_a(K_bp \lor \neg K_b\neg p)$, also equivalent to $K_aW_b$.
   For example, in $MC^3$ where B is muddy and C is clean, after the father’s announcement $\neg K_aW_BM_B$ holds.

7. ($\neg K_a/l_b$) ‘a is ignorant (of the fact) that b is ignorant about p’?
   $\neg K_a/l_bp := l_bp \land \neg K_al_bp$.
   $\equiv \neg K_bp \land \neg K_b\neg p \land \neg K_a(\neg K_bp \land \neg K_b\neg p)$.
   For example, in $MC^3$ where only A and B are muddy, after the father’s announcement $\neg K_a/l_BM_B$ holds. (For, B is ignorant whether he is muddy, but A consider it possible that A is clean, and then B would know that he is muddy.)

---

7 By analogy with Fine’s RI, one may also read $\neg K_a/l_bp$ as ‘a is Rumsfeld ignorant that b is ignorant about p’. However, the self-referential intent of the single-agent Rumsfeld ignorance is lost in this translation.
8. (\(\neg K_a/W_b\)) ‘a is ignorant (of the fact) that b knows whether p’: 
\[\neg K_a/W_b := W_b \land \neg K_a W_b\]
\[\equiv (K_b p \lor K_b \neg p) \land \neg K_a(K_b p \lor K_b \neg p).\]

For example, in MC\(^3\) where only A and B are muddy, after the father’s announcement \(\neg K_a/l_b M_B\) holds. (For, B is ignorant whether he is muddy, but A consider it possible that A is clean and hence B knows that he is muddy.)

9. (\(l_a K_b\)) ‘a is ignorant whether b knows p’: \(l_a K_b p \equiv \neg K_a K_b p \land \neg K_a \neg K_b p.\)

For instance, in the last MC\(^3\) example above, \(l_a K_b M_B\) holds, too.

Note that \(l_a K_b p\) implies that a considers p possible, i.e. the following is valid: 
\[\vdash l_a K_b p \rightarrow K_a p.\]
For, otherwise a would know that b could not know p, i.e. that \(\neg K_b p\) holds, which would contradict a’s ignorance about b’s knowledge of p.

Thus, our knowledge of a’s ignorance about b’s knowledge also gives us some knowledge on the factive compatibility of a’s knowledge.

10. (\(lK_{(a,b)}\)) ‘a is ignorant about b’s knowledge of p’:
\[lK_{(a,b)} := \neg K_a K_b p \land \neg K_a \neg K_b \neg p \equiv \neg K_a K_b p \land \neg K_a \neg K_b \neg p.\]

For example, in MC\(^3\), both \(lK_{(a,b)} M_A\) and \(lK_{(a,b)} M_B\) hold.

Note that, since \(l_a K_b p \equiv \neg K_a K_b p \land \neg K_a \neg K_b \neg p\) and \(\vdash \neg K_a K_b p \rightarrow \neg K_a \neg K_b \neg p\), we have 
\[\vdash l_a K_b p \rightarrow lK_{(a,b)} p.\]
The converse implication, however, is not valid. For example, in MC\(^3\), if B is muddy then (before the father’s announcement) \(lK_{(a,b)} M_B\) holds, but \(l_a K_b M_B\) does not hold.

11. (\(l_{(a,b)}\)) ‘a is ignorant whether b is ignorant about p’:
\[l_{(a,b)} := \neg K_a \neg K_b p \land \neg K_a \neg K_b \neg p \equiv \neg K_a \neg K_b p \land \neg K_a \neg K_b \neg p.\]

Note that, since \(\vdash K_a K_b p \rightarrow K_a \neg K_b p\) we have by contraposition that 
\[\vdash \neg K_a \neg K_b p \rightarrow \neg K_a \neg K_b p.\]
Therefore:
\[\vdash (\neg K_a \neg K_b p \land \neg K_a \neg K_b \neg p) \rightarrow (\neg K_a \neg K_b p \land \neg K_a \neg K_b \neg p), i.e., \vdash l_{(a,b)} p \rightarrow l_a K_b p.\]

Consequently, \(\vdash l_{(a,b)} p \rightarrow l_a K_b p\), i.e. the ignorance of a whether b is ignorant about p implies the factual ignorance of a whether p.

The converse implication \(l_a K_b p \rightarrow l_{(a,b)} p\), however, is not valid. For example, in MC\(^3\) where B is muddy, but C is not, after the father’s announcement \(l_a K_b M_B\) holds, whereas \(l_{(a,b)} M_B\) does not hold (because \(K_a \neg K_b \neg M_B\) holds).

Note further that \(l_{(a,b)} p \vdash lK_{(a,b)} p\), and therefore:
\[l_{(a,b)} p \equiv lK_{(a,b)} p \land l_{(a,b)} p\]
\[\equiv \neg K_a K_b p \land \neg K_a K_b \neg p \land \neg K_a \neg K_b \neg p \land \neg K_a \neg K_b \neg p\]
\[\equiv l_a K_b p \land l_a K_b \neg p.\]

The latest expression, which will be denoted as \(l_a (K_b) p\), says that a is completely ignorant about b’s knowledge about p. Thus, the complete ignorance of a about
b’s knowledge about p is equivalent to plain ignorance of a whether b is ignorant about p.

12. (i_b) ‘a is ignorant about b’s ignorance about p’: I_b\_b \_p.

For instance, in the last MC^3 example above, I_\_b \_b \_M holds, too. In fact, as we will note further, \(\neg K_a /\_b \_p \models I_b \_b \_p\). The converse implication is not valid, however. Indeed, again in MC^3, if only B is muddy, after the father’s announcement I_\_b \_b \_M holds but \(\neg K_a /\_b \_b \_M\) does not hold (for, then B knows that he is muddy).

Further, note the following chain of equivalences:
\[
I_b \_b \_p \equiv \neg K_a \_b \_p \land \neg K_a \_b \_p
\]
\[
\equiv \neg K_a (\neg K_b \_p \land \neg K_a \_p) \land \neg K_a (\neg K_b \_p \land \neg K_a \_p)
\]
\[
\equiv \neg (\neg K_a \_p \lor \neg K_a \_p) \land \neg K_a (\neg K_a \_p \lor \neg K_a \_p)
\]
\[
\equiv (\neg K_a \_b \_p \lor \neg K_a \_b \_p) \land \neg K_a (\neg K_a \_p \lor \neg K_a \_p)
\]
\[
\equiv (\neg K_a \_b \_p \lor \neg K_a \_b \_p) \land \neg K_a (\neg K_a \_p \lor \neg K_a \_p)
\]

Note that \(| \_b \_b \_p \rightarrow (\_b \_b \_p \lor \_b \_b \_p)\), i.e. \(| \_b \_b \_p \rightarrow IK_{a(b)}\_p\), hence \(| \_b \_b \_p \rightarrow IK_{a(b)}\_p\). The converse implication is not valid, which will follow from a stronger claim in the next proposition.

**Proposition 1** For any two agents, a and b, each with an S5 knowledge operator, the logical consequences and equivalences between the operators of relative ignorance, depicted on the figure below and listed beneath it, are valid:

\[
\neg K_a /\_b \models I_a \_b \_p \models IK_{a(b)} \models \neg K_a \_b \_p
\]

\[
\neg K_a /\_b \models I_a \_b \_b \models \neg K_a \_b \_b \models I_a \_b \_b \models \neg K_a \_b \_b
\]

- \(I_a (\neg K_b) \_p \equiv I_a \_b \_b \_p \models I_a \_b \_b \_p \models IK_{a(b)} \_b \_p \models \neg K_a \_b \_b \_p\).
- \(\neg K_a /\_b \_b \_p \models I_a \_b \_b \_p \models \neg K_a \_b \_b \_p \models \neg K_a \_b \_b \_p\).
- \(\neg K_a /\_b \_b \_p \models I_a \_b \_b \_p \models \neg K_a \_b \_b \_p\).
- \(\neg K_a /\_b \_b \_p \models I_a \_b \_b \_p \models \neg K_a \_b \_b \_p\).
- \(\neg K_a /\_b \_b \_p \models I_a \_b \_b \_p \models \neg K_a \_b \_b \_p\).
- \(I_a \_b \_b \_p \models IK_{a(b)} \_b \_p\).

126
On relative ignorance

Furthermore, only the logical consequences following from those listed above by transitivity of \(\models\) are valid.

Proof. All valid consequences and some of the non-valid ones have already been proved or are straightforward. For all non-valid consequences, simple abstract countermodels are provided in the Appendix, and almost all (even, possibly all) of these consequences can be refuted in variants of the 3 Muddy Children scenario. For example:

1. To show \(\text{ll}_{\{a,b\}} p \not\models \text{ll}_b p\), consider \(\text{MC}^1\) where \(A\) does not know whether she is muddy or not, but she knows that \(B\) knows whether she is muddy or not and she can imagine either case, so \(\text{ll}_{\{a,b\}} \text{M}_A\) holds, but \(\text{ll}_b \text{M}_A\) does not hold.

2. To show \(\text{ll}_b p \not\models \text{ll}_b \text{K}_b p\), consider again \(\text{MC}^3\) where \(C\) is not muddy. Then, after the father’s announcement that at least one child is muddy, \(A\) can imagine that \(B\) knows that he is muddy (if \(A\) is not muddy) and can also imagine that \(B\) is ignorant whether he is muddy or not (if \(A\) is muddy). Thus, \(\text{ll}_b \text{M}_B\) holds, but \(\text{ll}_b \text{K}_B \text{M}_B\) does not hold (because \(A\) knows that \(B\) does not know that he is not muddy, i.e., \(\text{K}_A \text{K}_B \text{M}_B\) holds).

\[\]

4 Iterated relative ignorance

Let us now look at the iterated versions of the relative ignorance operators identified in the previous section. For each such operator \(O\), I define \(O^np := O\ldots n\text{times}...O p\).

For instance: \((\neg \text{K}_a \neg \text{K}_b)^2 p = \neg \text{K}_a \neg \text{K}_b \neg \text{K}_a \neg \text{K}_b p\); \((\text{ll}_{\{a,b\}})^3 p = \text{ll}_{\{a,b\}} \text{ll}_{\{a,b\}} \text{ll}_{\{a,b\}} p\), etc.

Proposition 2 Given two agents, \(a\) and \(b\), each with an \(S5\) knowledge operator, the following hold.

1. \((\neg \text{K}_a \text{K}_b)^2 p \not\models \neg \text{K}_a \text{K}_b p\), \(\neg \text{K}_a \text{K}_b p \not\models \neg \text{K}_a \text{K}_b^2 p\).
   \((\neg \text{K}_a \text{K}_b)^3 p \models \neg \text{K}_a \text{K}_b p\), \(\neg \text{K}_a \text{K}_b p \not\models \neg \text{K}_a \text{K}_b^3 p\).
   \((\neg \text{K}_a \text{K}_b)^4 p \not\models \neg \text{K}_a \text{K}_b p\), \(\neg \text{K}_a \text{K}_b p \not\models \neg \text{K}_a \text{K}_b^4 p\).
   \((\neg \text{K}_a \text{K}_b)^5 \equiv (\neg \text{K}_a \text{K}_b)^2 p\).

2. \((\neg \text{K}_a \neg \text{K}_b)^2 p \models \neg \text{K}_a \neg \text{K}_b p\).

3. \((\neg \text{K}_a/\text{K}_b)^2 p \models \neg \text{K}_a/\text{K}_b p\), \(\neg \text{K}_a/\text{K}_b p \not\models (\neg \text{K}_a/\text{K}_b)^2 p\).
   \((\neg \text{K}_a/\text{K}_b)^3 p \equiv (\neg \text{K}_a/\text{K}_b)^2 p\).

4. \(\neg \text{K}_a/\neg \text{K}_b p\), \(\neg \text{K}_a/\neg \text{K}_b^2 p\), \(\neg \text{K}_a/\neg \text{K}_b^3 p\), \(\neg \text{K}_a/\neg \text{K}_b^4 p\), \(\neg \text{K}_a/\neg \text{K}_b^5 p\) are pairwise incomparable w.r.t. logical consequence.

5. \(\neg \text{K}_a b p \models (\neg \text{K}_a b)^2 p\), \((\neg \text{K}_a b)^2 p \not\models \neg \text{K}_a b p\).
   \((\neg \text{K}_a b)^3 p \equiv (\neg \text{K}_a b)^2 p\).
Conjecture 3 For every \( m \), the following hold in terms of logical consequence:

1. \((-K_a/W_b)^p \models (-K_a/W_b)^p\) for any \( m \)
2. \((-K_a/W_b)^p \models (-K_a/W_b)^p\) for any \( m \)
3. \((-K_a/W_b)^p \models (-K_a/W_b)^p\) for any \( m \)
4. \((-K_a/W_b)^p \models (-K_a/W_b)^p\) for any \( m \)

Proof. Proving all valid consequences is a routine, but long and laborious exercise. Likewise for refuting the non-valid ones. Instead, all these have been verified with the web-based Modal Logic Tableau Prover MOLTAP [35]. Some of the refuting counter-models, generated by MOLTAP, are provided in the appendix.

A number of questions regarding the precise behaviour of the hierarchies of some of the operators of relative ignorance are left unsettled, and some of them are listed as conjectures below.

Conjecture 3 For every \( m, n \in \mathbb{N} \), such that \( m \neq n \), the following hold in terms of logical consequence:

1. \((-K_a/W_b)^p \models (-K_a/W_b)^p\)
2. \((-K_a/W_b)^p \models (-K_a/W_b)^p\)
3. \((-K_a/W_b)^p \models (-K_a/W_b)^p\)
4. \((-K_a/W_b)^p \models (-K_a/W_b)^p\)

Conjecture 4 For every \( m, n \in \mathbb{N} \):

1. If \((m, n) \neq (1, 1)\) then \((l_a b)^n p \models (l_a b)^n p\)
2. \((l_a b)^n p \models (l_a b)^n p\)
5 Concluding remarks

The formal study of multi-agent ignorance is still in its very early stage. The aim of this paper was modest: to undertake a systematic study of the various natural patterns of relative ignorance and to establish a range of technical results describing their comparative behaviour and relationships. A number of both conceptual and technical issues are yet to be studied, including mutual ignorance, higher-order (iterated) mutual ignorance, and common ignorance. These are left for future exploration. In conclusion:

while already quite knowledgeable about knowledge,
we are still quite ignorant about ignorance.

Acknowledgements I thank Twan van Laarhoven for his help with running the web version of the tool MOLTAP, implemented by him. I also thank the anonymous reviewer for some corrections and references.

Appendix: some proof details and countermodels

Counter-models for the proof of Proposition 1. Here are 10 models (generated with MOLTAP) which suffice to refute all non-valid consequences stated in Proposition 1.
For convenience, here is the list of the 12 operators of relative ignorance listed in Section 3:

1. $\neg K_a K_b$
2. $\neg K_a \neg K_b$
3. $\neg K_a / K_b p$
4. $\neg K_a / \neg K_b p$
5. $\neg K_a W b$
6. $\neg K_a I b$
7. $I K_b$
8. $l a K_b$
9. $I K_a(K_b)$
10. $l a l_b$
11. $I K_a(l_b)$
12. $l a I b$

I leave it to the reader to check that every non-valid consequence, as stated in Proposition 1, is falsified in the designated (double-circled) state of at least one of the models above. Most of these claims for each of the respective falsifying models are listed below, using the numeration of the operators above, where e.g. "$2 \not|= 6$" means $\neg K_a \neg K_b p \not|= \neg K_a W b p$, etc.

$M_1$: $2 \not|= 6, 2 \not|= 7, 2 \not|= 9, 2 \not|= 10, 2 \not|= 12, 5 \not|= 1, 5 \not|= 4, 5 \not|= 6, 5 \not|= 7, 5 \not|= 10, 9 \not|= 7, 10 \not|= 4, 10 \not|= 6, 5 \not|= 11, 5 \not|= 12, 11 \not|= 4, 11 \not|= 12$.

$M_2$: $1 \not|= 10, 1 \not|= 11$.

$M_3$: $3 \not|= 4, 3 \not|= 6, 3 \not|= 7, 3 \not|= 8, 3 \not|= 12, 9 \not|= 7, 10 \not|= 6, 11 \not|= 6, 11 \not|= 12$.

$M_4$: $2 \not|= 3, 2 \not|= 8, 4 \not|= 3, 4 \not|= 6, 5 \not|= 8, 9 \not|= 3, 9 \not|= 8, 11 \not|= 3, 1 \not|= 8$.

$M_5$: $1 \not|= 5, 1 \not|= 8, 6 \not|= 5, 6 \not|= 8, 6 \not|= 12, 10 \not|= 5, 10 \not|= 8$.

$M_6$: $2 \not|= 7, 2 \not|= 4, 3 \not|= 6, 8 \not|= 8, 8 \not|= 8, 8 \not|= 11, 11 \not|= 7, 12 \not|= 4, 12 \not|= 7, 12 \not|= 11$.

$M_7$: $6 \not|= 2, 6 \not|= 3, 6 \not|= 9, 6 \not|= 11, 8 \not|= 2, 8 \not|= 3, 8 \not|= 9, 10 \not|= 2, 10 \not|= 9, 10 \not|= 11, 12 \not|= 2, 12 \not|= 9$.

$M_8$: $4 \not|= 11, 7 \not|= 3, 7 \not|= 8, 9 \not|= 11, 12 \not|= 8$.

$M_9$: $7 \not|= 2, 7 \not|= 4, 7 \not|= 9$.

$M_{10}$: $12 \not|= 3$. 

130
Counter-models for the proof of Proposition 2. Here are a few models, generated by MOLTAP, which refute some of the non-valid consequences stated in Proposition 2.

The notation that I use for the models indicates their purpose here as follows: for an operator $O$ appearing under number $n$ in the list, a model refuting the consequence $O^p \models O^m p$ will be denoted by $M_{n:k\rightarrow m}$.

- $M_{9:1\rightarrow2}$: $I_aK_b p \not\models (I_aK_b)^2 p$
- $M_{10:1\rightarrow2}$: $IK_{(a;b)} p \not\models IK_{(a;b)}^2 p$
- $M_{10:1\rightarrow3}$: $IK_{(a;b)} p \not\models IK_{(a;b)}^3 p$
- $M_{10:1\rightarrow4}$: $IK_{(a;b)} p \not\models IK_{(a;b)}^4 p$

- $M_{11:1\rightarrow2}$: $II_{(a;b)} p \not\models II_{(a;b)}^2 p$

- $M_{2:1\rightarrow2}$: $\neg K_a/\neg K_b p \not\models (\neg K_a/\neg K_b)^2 p$

- $M_{3:1\rightarrow3}$: $(\neg K_a/\neg K_b)p \not\models (\neg K_a(\neg K_b))^3 p$

- $M_{5:2\rightarrow1}$: $(\neg K_a b)^2 p \not\models (\neg K_a b)p$

Because of an apparent minor bug in the tool, some of these contain isolated worlds that are to be inserted in place of the worlds with empty labels.
\begin{itemize}
  \item $\mathcal{M}_{8,1+2}^\cdot: \neg \text{K}_a / \text{W}_b \neg p \not\models (\neg \text{K}_a / \text{W}_b)^2 p$
  \item $\mathcal{M}_{8,1+3}^\cdot: \neg \text{K}_a / \text{W}_b \neg p \not\models (\neg \text{K}_a / \text{W}_b)^3 p$
  \item $\mathcal{M}_{7,1+2}^\cdot: \neg \text{K}_a / \text{I}_b p \not\models (\neg \text{K}_a / \text{I}_b)^2 p$
  \item $\mathcal{M}_{4,3+2}^\cdot: (\neg \text{K}_a / \neg \text{K}_b)^3 p \not\models (\neg \text{K}_a / \neg \text{K}_b)^2 p$
  \item $\mathcal{M}_{4,3+3}^\cdot: (\neg \text{K}_a / \neg \text{K}_b)^4 p \not\models (\neg \text{K}_a / \neg \text{K}_b)^3 p$
  \item $\mathcal{M}_{10,2+1}^\cdot: \text{I}_K^2 p_{(a,b)} / \neg p \not\models \text{I}_K^2 p_{(a,b)}$
\end{itemize}
On relative ignorance

$\mathcal{M}_{4,3;4}^1: (\neg K_a / \neg K_b)^3 p \not\models (\neg K_a (\neg K_b))^3 p$

$\mathcal{M}_{12,1;3}^1: l_{a,b} p \not\models (l_{a,b})^3 p$

$\mathcal{M}_{12,1;3}^2: l_{a,b} p \not\models (l_{a,b})^2 p$

$\mathcal{M}_{12,1;4}^1: l_{a,b} p \not\models (l_{a,b})^4 p$

$\mathcal{M}_{12,1;4}^2: l_{a,b} p \not\models (l_{a,b})^3 p$

$\mathcal{M}_{10,5;3}^1: I_{(a,b)}^5 p \not\models I_{(a,b)}^3 p$

$\mathcal{M}_{10,5;3}^2: I_{(a,b)}^5 p \not\models I_{(a,b)}^3 p$

$\mathcal{M}_{12,2;3}^1: l_{a,b} p \not\models (l_{a,b})^2 p$

$\mathcal{M}_{12,2;3}^2: l_{a,b} p \not\models (l_{a,b})^3 p$

$\mathcal{M}_{12,2;3}^3: l_{a,b} p \not\models (l_{a,b})^3 p$
\[ M_{7,2+3} : ~ (\neg K_a / b)^2 p \not\models (\neg K_a / b)^3 p \]

\[ M_{12,2+4} : ~ (l_a b)^3 p \not\models l_a b p \]

\[ M_{12,3+4} : ~ (l_a b)^3 p \not\models l_a b p \]

\[ M_{9,1+4} : ~ I_a K_b p \not\models I_a K_b p \]

\[ M_{6,3+4} : ~ (\neg K_a / \neg K_b)^3 p \not\models (\neg K_a (\neg K_b)) p \]

\[ M_{4,4+3} : ~ (\neg K_a / \neg K_b)^3 p \not\models (\neg K_a (\neg K_b))^2 p \]
On relative ignorance

\( \neg K_a W_b p, (\neg K_a W_b)^2 p, (\neg K_a W_b)^3 p, (\neg K_a W_b)^4 p, (\neg K_a W_b)^5 p \) are pairwise incomparable w.r.t. logical consequence:

- \( \mathcal{M}_{6;2\rightarrow 1} : (\neg K_a W_b)^2 p \not\models \neg K_a W_b p \),
- \( \mathcal{M}_{6;2\rightarrow 3} : (\neg K_a W_b)^2 p \not\models (\neg K_a W_b)^3 p \),
- \( \mathcal{M}_{6;4\rightarrow 3} : (\neg K_a W_b)^4 p \not\models (\neg K_a W_b)^3 p \),
- \( \mathcal{M}_{6;4\rightarrow 5} : (\neg K_a W_b)^4 p \not\models (\neg K_a W_b)^5 p \),
- \( \mathcal{M}_{6;3\rightarrow 1} : (\neg K_a W_b)^3 p \not\models \neg K_a W_b p \),
- \( \mathcal{M}_{6;3\rightarrow 2} : (\neg K_a W_b)^3 p \not\models (\neg K_a W_b)^2 p \),
- \( \mathcal{M}_{10;3\rightarrow 1} : I K_4^4 (a;b)p \not\models I K_4^4 (a;b)p \),
- \( \mathcal{M}_{10;3\rightarrow 4} : I K_4^4 (a;b)p \not\models (\neg K_a W_b)^4 p \),
- \( \mathcal{M}_{10;3\rightarrow 5} : I K_4^4 (a;b)p \not\models (\neg K_a W_b)^5 p \),
- \( \mathcal{M}_{10;5\rightarrow 4} : I K_4^4 (a;b)p \not\models (\neg K_a W_b)^4 p \),
\[ M_{6.4\rightarrow2}: (\neg K_a W b) p \not\models (\neg K_a W b)^2 p \]
\[ M_{7.2\rightarrow1}: (\neg K_a / b) p \not\models (\neg K_a / b)^2 p, \]
\[ M_{10.3\rightarrow5}: IK_{(a,b)}^3 p \not\models IK_{(a,b)}^5 p \]
\[ M_{10.4\rightarrow2}: IK_{(a,b)}^4 p \not\models IK_{(a,b)}^2 p \]

\[ M_{10.2\rightarrow3}: IK_{(a,b)}^2 p \not\models IK_{(a,b)}^3 p \]
\[ M_{10.2\rightarrow5}: IK_{(a,b)}^2 p \not\models IK_{(a,b)}^5 p \]
\[ M_{10.2\rightarrow4}: IK_{(a,b)}^2 p \not\models IK_{(a,b)}^4 p \]
• $M_{10:3-1}: 1K^3_{(a,b)}p \not\models 1K_{(a,b)}p$
• $M_{10:4-3}: 1K^4_{(a,b)}p \not\models 1K^3$

• $M_{10:3-2}: 1K^3_{(a,b)}p \not\models 1K^2_{(a,b)}p$
• $M_{10:4-5}: 1K^4_{(a,b)}p \not\models 1K^5_{(a,b)}p$

• $M_{12:3-2}: (l_{a,b})^3 p \not\models (l_{a,b})^2 p$
• $M_{12:4-1}: (l_{a,b})^4 p \not\models l_{a,b}p$
References


On relative ignorance


Valentin Goranko


Valentin Goranko
Department of Philosophy, Stockholm University
Email: valentin.goranko@philosophy.su.se