Economics & Philosophy (2023), 1–22
doi:10.1017/S0266267123000238

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Reasoning with reasons: Lewis on common knowledge

Huub Vromen

Radboud University, Department of Philosophy, Erasmusplein 1, 6525 HT Nijmegen, Netherlands
Email: hjvromen@icloud.com

(Received 26 August 2022; revised 20 March 2023; accepted 02 May 2023)

Abstract
David Lewis is widely regarded as the philosopher who introduced the concept of common knowledge. His account of common knowledge differs greatly from most later accounts in philosophy and economy, with the central notion of his theory being ‘having reason to believe’ rather than ‘knowledge’. Unfortunately, Lewis’s account is rather informal, and the argument has a few gaps. This paper assesses two major attempts to formalise Lewis’s account and argues that these formalisations are missing a crucial aspect of this account. Therefore, a new reconstruction is proposed, which explicitly discusses ‘reasons’ and uses a logic inspired by justification logic.

Keywords: Common knowledge; reasons; rationality; belief; logic

1. Introduction

Common knowledge plays an important role in our social interactions, communication, and coordination of activities. For instance, Clark and Marshall (1981) show that the felicitous use of a definite referring expression requires common knowledge, and Fagin *et al.* (1995) explain how common knowledge enables people to coordinate their actions.

Logicians, economists, philosophers, and psychologists have analysed the concept of common knowledge, and various, but broadly equivalent, formalisations of common knowledge have been proposed (Hoek and Meyer 1997; Heifetz 1999; van Ditmarsch *et al.* 2008). Common knowledge is a stronger notion than shared knowledge. Ann and Bob have shared knowledge about φ if and only if Ann knows φ and Bob knows φ. Common knowledge requires not only shared knowledge but also that Ann and Bob both know that both know φ, Bob and Ann both know that both know that both know φ, and so on ad infinitum.

Intuitively, common knowledge is information that is public and fully transparent, for instance, after a public announcement. Common knowledge also ensues from situations like the following:

> Suppose that you and I are dining together and that we are seated across from one another and that on the table between us is a rather conspicuous candle. We would therefore be in a situation in which I am facing the candle and you, and you are facing the candle and me. (Consequently, a situation in which S is facing the candle and A, who is facing the candle and S, who is facing the candle and A, who is facing the candle and S, who is facing ...) I submit that were this situation to be realized, you and I would mutually know* that there is a candle on the table.[1] (Schiffer 1972: 31)

> A goat walks into the room, or all of the lights suddenly go out. In such a case, it immediately becomes common knowledge that the event has happened - that there is a goat in the room, or that the lights have gone out. (Stalnaker 2014: 47)

The concept of common knowledge has been criticised for being unrealistic. The basic problem is that knowledge is something psychological, a mental state. Thus, common knowledge requires that individuals can know the mental states of other individuals, which is impossible according to some philosophers (Paternotte 2017: 452). But more importantly, common knowledge presupposes the possibility of having an infinite series of mental states about mental states about mental states ..., which is arguably too much to fit in our minds (Sperber and Wilson 1995).

Given that common knowledge seems to be indispensable as well as problematic, it is worth looking again at the origins of the concept. Although David Lewis (1969) is widely regarded as the philosopher who introduced the concept of common knowledge, his account of common knowledge differs greatly from most later accounts. The central notions of his theory are not psychological states, like knowledge or belief, but 'reasons to believe'. Reasons to believe are normative relations between persons and propositions, not mental states.

Lewis discusses what essential requirements a situation must meet in order to generate higher-order reasons to believe. Intuitively, these requirements are that the situation is open and transparent for all participants. A situation that satisfies these requirements Lewis calls a 'basis for common knowledge'. He proves that such a basis for common knowledge indeed generates an infinite series of higher-order reasons to believe: everyone has reason to believe, everyone has reason to believe that everyone has reason to believe, and so on.

Unlike most theories about common knowledge, Lewis's account does not suffer from the problems associated with higher-order mental states. That is why his theory deserves renewed attention. Unfortunately, his account is rather informal, and the argument is incomplete. There have been several attempts to formalise his account and close the gaps, but none of these proposals is entirely satisfactory.

[1]Schiffer uses the term 'mutual knowledge*' instead of 'common knowledge'.

https://doi.org/10.1017/S0266267123000238 Published online by Cambridge University Press

This paper aims to give a new interpretation of Lewis's theory that closes the gaps in his account, using logical tools that have become available in the last decade.

In section 2, I examine and criticise Lewis's account, starting with the central concepts of his theory: 'reasons to believe', and 'basis for common knowledge'. Then I discuss Lewis's theorem which says that a basis for common knowledge generates an infinite sequence of higher-order reasons to believe. It will be shown that there is a gap in his argument for this thesis. Several philosophers have tried to close this gap, either by making additional assumptions or by taking a different approach to proving the theorem.

Section 3 discusses how Cubitt and Sugden (2003) seek to close the gap in Lewis's argument using an interpretation of the core concepts 'reason to believe' and 'indication'. This discussion is followed by a presentation of Sillari's (2005) approach, which uses a possible worlds semantics to implement Lewis's theory.

Section 4 proposes an alternative way of completing Lewis's argument. I give a novel interpretation of the concept of reason to believe that is more in line with Lewis's text, using a logic that makes it possible to reason about reasons. In this logic, which is motivated on independent grounds, Lewis's argument becomes fully explicit and valid.

While the main focus of the paper is on reasons to believe, actual beliefs, too, are important in Lewis's theory of conventions, because people act on their beliefs when solving coordination problems. Lewis gives a rather informal account of the relation between beliefs and reasons to believe; section 5 discusses proposals for making his account more explicit.

2. Lewis's account of common knowledge

In this section, I outline Lewis's theory of common knowledge. His definition goes as follows:

> Let us say that it is *common knowledge* in a population *P* that ___ if and only if some state of affairs *A* holds such that:
>
> (1) Everyone in *P* has reason to believe that *A* holds.
> (2) *A* indicates to everyone in *P* that everyone in *P* has reason to believe that *A* holds.
> (3) *A* indicates to everyone in *P* that ___.
>
> We can call any such state of affairs *A* a *basis* for common knowledge in *P* that ___. (Lewis 1969: 56).

According to Lewis, a basis for common knowledge generates an infinite chain of higher-order reasons to believe: everyone in *P* has reason to believe that ___, everyone in *P* has reason to believe that everyone in *P* has reason to believe ___, and so on ad infinitum. Let us call this 'Lewis's theorem'.

https://doi.org/10.1017/S0266267123000238 Published online by Cambridge University Press

Note that Lewis's theorem does not say or imply that a basis for common knowledge generates an infinite series of knowledge states. Hence, the term 'common knowledge' is a misnomer, as Lewis himself came to realise:

> A better definition of overt belief, under the name of 'common knowledge', may be found in my *Convention* (Cambridge, Mass., 1969), pp. 52-60. That name was unfortunate since there is no assurance that it will be knowledge, or even that it will be true. (Lewis 1978: 44, fn13)

Therefore, it would be more appropriate to call this infinite sequence of reasons to believe 'common reason to believe' instead of 'common knowledge'. Commentators on Lewis's theory usually keep the name 'common knowledge' (Clark and Marshall 1981; Vanderschraaf 1998; Cubitt and Sugden 2003), but some use the term 'common reason to believe' (Sillari 2005). In the following both terms will be used, conforming to the authors discussed.

Although 'reason to believe' and 'indication' are central to Lewis's definition of common knowledge, he does not provide formal definitions of these notions. In the following, I will try to provide definitions that stay as close as possible to Lewis's text. An obvious starting point for defining 'reason to believe' is to treat it as a relation between individuals and propositions. This is the approach taken by philosophers who have analysed Lewis's argument (Vanderschraaf 1998; Cubitt and Sugden 2003; Sillari 2005). I will discuss this approach in section 4 and then proceed to offer a novel approach, based on the idea that reasons to believe are objects of a sort, at least in the sense that they can be quantified over. For now, I will remain neutral on this point and use $R_i\varphi$ as an abbreviation for 'i has reason to believe φ'.[2]

'Indication' is defined by Lewis as follows:

> Let us say that A indicates to someone i that ___if and only if, if i had reason to believe that A held, i would *thereby* have reason to believe that ___ (Lewis 1969: 52-53, my emphasis).

Following Vanderschraaf (1998), Cubitt and Sugden (2003) and Sillari (2005), I treat 'indication' as a relation between an individual, a state of affairs, and a proposition, using $Ind_i\, A\varphi$ as an abbreviation for 'the state of affairs A indicates to agent i that φ'. The question now is how we should interpret this relation. The difficulty is that Lewis uses the subjunctive mood and 'if ... thereby' instead of the standard logical operator 'if ... then'. This indicates that he does not mean a material implication. Moreover, the assumption that a material implication is meant leads to implausible conclusions. For, assume:

$$Ind_i\, A\varphi \leftrightarrow (R_i(A \text{ holds}) \rightarrow R_i\varphi). \tag{A0}$$

[2]I will use i and j for individuals, in accordance with the standard notation in the literature and not, like Lewis, x and y.

https://doi.org/10.1017/S0266267123000238 Published online by Cambridge University Press

The left-to-right part of (A0) states

$$Ind_i\, A\varphi \to (R_i(A \text{ holds}) \to R_i\varphi)$$

which is equivalent to

$$(Ind_i\, A\varphi \wedge R_i(A \text{ holds})) \to R_i\varphi \tag{A1}$$

This seems intuitively plausible, and most importantly, it captures Lewis's statement: "Consider next that our definition of indication yields a principle of detachment: if A indicates to x that ___ and x has reason to believe that A holds, then x has reason to believe that ___." (Lewis 1969: 53). Lewis uses this inference in every step of his argument.

However, on the face of it, at least, the right-to-left part of (A0)

$$(R_i(A \text{ holds}) \to R_i\varphi) \to Ind_i\, A\varphi$$

does not seem correct. For one thing, it makes $Ind_i\, A\varphi$ vacuously true if i does not have a reason to believe A. However, in Lewis's analysis, the indication relation is only used in contexts in which an individual does have reason to believe A, and therefore, this objection does not carry much weight. A more serious objection is that, according to the right-to-left part of (A0), $Ind_i\, A\varphi$ is true for every proposition φ such that i has reason to believe φ, which makes the indication relation trivial.

Apparently, the indication relation must be stronger than material implication. In section 4, I propose a new account of indication that does better justice to Lewis's formulation.

2.1. Lewis's theorem

We can now begin to explicate Lewis's definition of a basis for common knowledge as follows: a state of affairs A is a basis for common knowledge of a proposition φ in a group P if and only if A holds and the following hold for all i and j in P:

$$R_i(A \text{ holds}) \tag{C1}$$

$$Ind_i\, A\big(R_j(A \text{ holds})\big) \tag{C2}$$

$$Ind_i\, A\varphi \tag{C3}$$

According to Lewis, these conditions are necessary but not sufficient for generating an infinite chain of higher-order reasons to believe. In addition, "suitable ancillary premises regarding our rationality, inductive standards, and background information" (Lewis 1969: 53) are required. Lewis develops these premises as follows.

> Consider that if A indicates something to i, and if j shares i's inductive standards and background information, then A must indicate the same thing to j. (53)

because

> What A indicates to i will depend, therefore, on i's inductive standards and background information. (53)

https://doi.org/10.1017/S0266267123000238 Published online by Cambridge University Press

Lewis then states:

> Therefore, if A indicates to i that j has reason to believe that A holds, and if A indicates to i that ___, and if i has reason to believe that j shares i's inductive standards and background information, then A indicates to i that j has reason to believe that ___ (this reason being j's reason to believe that A holds). (53)[3]

Accordingly, Lewis assumes that the members of P have reason to believe they share the same inductive standards and background information, at least nearly enough so that A will indicate the same things to all of them and, on that basis, he assumes the following:

$$Ind_i\, A\big(R_j(A \text{ holds})\big) \wedge Ind_i\, A\varphi \rightarrow Ind_i\, A\big(R_j\varphi\big) \tag{S}$$

However, Lewis offers no formal proof that this follows from his definition of indication, or from his description of rationality, inductive standards and background information. Therefore, we have to treat this as an axiom. This axiom is needed to introduce higher-order indications that, using (A1), then give rise to higher-order reasons to believe.

Lewis now uses (S) to outline a proof that, if conditions (C1)–(C3) are met, there is common knowledge of φ, i.e. that all finitely nested formulas $R_iR_j \dots R_k\varphi$ hold for all $i, j, \dots, k$ in P (Lewis's theorem). The first steps of his proof are these:

	— first-order reasons to believe	
(1)	$\forall i \in P : Ind_i\, A\varphi$	C3
(2)	$\forall i \in P : R_i\varphi$	C1, 1, A1
	— second-order reasons to believe	
(3)	$\forall i, j \in P : Ind_i\, A\big(R_j\varphi\big)$	C2, 1, S
(4)	$\forall i, j \in P : R_iR_j\varphi$	C1, 3, A1
	— third-order reasons to believe	
(5)	$\forall i, j, k \in P : Ind_i\, A\big(R_jR_k\varphi\big)$	C2, 3, S
(6)	$\forall i, j, k \in P : R_iR_jR_k\varphi$	C1, 5, A1
	— higher-order reasons to believe	
(7)	etc.	

In the next section, I will first explain how Cubitt and Sugden (2003) formalise this account and thereby clarify the plausibility of (S) and how Sillari (2005) proposes a very different approach to prove the theorem.

[3]Remarkably, on p. 54 Lewis uses the expression 'I believe that you share my inductive standards and background information' instead of 'I have reason to believe that you share my inductive standards and background information'. This is probably a slip of the pen, because otherwise his account is only about reasons to believe and not about actual beliefs.

3. Previous proposals for developing the theory

3.1. Cubitt and Sugden

Cubitt and Sugden (2003) start their reconstruction of Lewis's theory by giving an interpretation of the key concepts of this theory. First, 'having a reason to believe' is interpreted as a two-place relation between agents and propositions, as follows:

> To say that some individual i has reason to believe some proposition x is to say that x is true within some logic of reasoning that is *endorsed* by (that is, accepted as a normative standard by) person i. For x to be true within such a logic of reasoning, it must *either* be treated as self-evident *or* be derivable from propositions that are treated as self-evident using the inference rules of the logic. Self-evidence may be either *a priori* or obtained through observation; the rules of inference may be deductive or inductive. (Cubitt and Sugden 2003: 184)

Second, 'indication' is interpreted as a three-place relation between agents, states of affairs and propositions. Lewis's formulation "if i had reason to believe that A held, i would *thereby* have reason to believe that x" is read as "i's reason to believe that A holds *provides* i's reason for believing that x is true". When combined with the above definition of reason to believe, this leads to:

> Our interpretation of the formula 'A ind_i x' is that, in the logic of reasoning that i endorses, there is an inference rule which legitimates inferring x from 'A holds'. (187)

The central concept in this interpretation of Lewis's account is thus the logic of reasoning that someone endorses. This idea has been further elaborated in Cubitt and Sugden (2014). In this paper, the phrase 'logic of reasoning' is replaced by 'reasoning scheme'. Each agent has his own reasoning scheme, which consists of a non-empty set of axioms and a set of inference rules that contains all rules of logically valid inferences. This reasoning scheme is taken to be consistent, i.e. no contradiction can be derived.

On this basis they assume that the indication relation has the following properties:[4]

$$\forall i \in P : (Ind_i\, A\varphi \wedge R_i(A \text{ holds})) \rightarrow R_i\varphi \tag{A1}$$

$$\forall i, j \in P : Ind_i\, A\, \psi \rightarrow R_i\big(Ind_j\, A\, \psi\big) \tag{C4}$$

$$\forall i, j \in P : \big(Ind_i\, A\big(R_j(A \text{ holds})\big) \wedge R_i\big(Ind_j\, A\varphi\big)\big) \rightarrow Ind_i\, A\big(R_j\varphi\big) \tag{A6}$$

(A1) is implied by the above definition of indication. It is just the left-to-right part of Lewis's definition of indication, discussed in section 2.

[4]In fact, Cubitt and Sugden (2003) assume a stronger version of (A6):

$$\forall i, j \in P : \big(Ind_i\, A\big(R_jA'\big) \wedge R_i\big(Ind_j\, A'\, x\big)\big) \rightarrow Ind_i\, A\big(R_jx\big)$$

But in their proof, they only use the special case in which $A'=A$.

https://doi.org/10.1017/S0266267123000238 Published online by Cambridge University Press

(C4) is based on Lewis's assumption "Suppose you and I do have reason to believe we share the same inductive standards and background information, at least nearly enough so that A will indicate the same things to both of us ..." (Lewis 1969: 53). This axiom is not required to hold for all possible propositions, but only for the proposition φ in the definition of a basis for common knowledge and for propositions that express first-order or higher-order reasons to believe φ.

(A6) is motivated as follows:

> (A6) says that if i's logic legitimates an inference from 'A holds' to the proposition that j has reason to believe 'A′ holds', and if i has reason to believe that j's logic legitimates an inference from 'A′ holds' to x, then i's logic legitimates an inference from 'A holds' to the proposition that j has reason to believe x.' (Cubitt and Sugden 2003: 188)

As a further explanation they add

> If i has reason to believe that j's logic legitimates an inference from (A′ holds) to x, then, because i's logic obeys the rules of deductive inference, i has reason to believe [R_j (A' holds) $\rightarrow R_j(x)$]. Hence, given the antecedent of the previous sentence, if i's logic allows an inference from (A holds) to R_j (A′ holds), it also allows an inference from (A holds) to R_j (x). Notice that A6 attributes deductive inference only to i's reasons to believe. It does not postulate anything about what i has reason to believe about the inference rules endorsed by j. (188, fn17)

Interestingly, Lewis's axiom (S) deductively follows from (C4) and (A6), as is easy to see.

These three assumptions suffice to prove Lewis's theorem. The first steps of the proof in this reconstruction are as follows:

	— first-order reasons to believe	
(1)	$\forall i \in P : Ind_i\, A\varphi$	C3
(2)	$\forall i \in P : R_i\varphi$	C1, 1, A1
	— second-order reasons to believe	
(3)	$\forall i,j \in P : R_i\left(Ind_j\, A\varphi\right)$	1, C4
(4)	$\forall i,j \in P : Ind_i\, A\left(R_j\varphi\right)$	C2, 1, A6
(5)	$\forall i,j \in P : R_iR_j\varphi$	C1, 4, A1
	— third-order reasons to believe	
(6)	$\forall i,j,k \in P : R_i\left(Ind_j\, A(R_k\varphi)\right)$	5, C4
(7)	$\forall i,j,k \in P : Ind_i\, A\left(R_jR_k\varphi\right)$	C2, 5, A6
(8)	$\forall i,j,k \in P : R_iR_jR_k\varphi$	C1, 5, A1
	— higher-order reasons to believe	
(9)	etc.	

Cubitt and Sugden have clarified Lewis's theory by offering an interpretation of the two basic concepts of 'reason to believe' and 'indication'. These interpretations are based upon a model in which each person has a logic of reasoning with axioms and inference rules. In addition, they have replaced the crucial axiom (S) by two axioms, both of which, with the help of the aforementioned interpretations, they have argued to be plausible. These two axioms together imply Lewis's axiom (S). However, Cubitt and Sugden's interpretation of 'has reason to believe' has some drawbacks.

First, it offers no way of representing that someone may have reasons for believing two contradictory propositions. If an agent i has reason to believe a proposition φ then φ is a theorem in the reasoning scheme that i endorses, according to Cubitt and Sugden (2014). But φ and not-φ cannot both be true in i's reasoning scheme because a reasoning scheme is supposed to be consistent. It is, however, quite conceivable, that someone has reasons to believe two contradictory propositions. Take, for example, the Nixon diamond scenario (Horty 2012: 27-28). The example is set in the time when Nixon was president of the US. The fact that Nixon is a Quaker and that Quakers usually are pacifists is a reason to believe that Nixon is a pacifist. At the same time, the fact that Nixon is a Republican and that Republicans tend not to be pacifists is a reason to believe that Nixon is not a pacifist.

Second, Cubitt and Sugden's account implies that every agent has reason to believe every tautology because his reasoning scheme contains all rules of valid inferences. This consequence is not implausible in itself but it is unclear whether it is in the spirit of Lewis's approach, according to which our actual beliefs are governed by our reasons to believe: "Anyone who has reason to believe something will come to believe it, provided he has a sufficient degree of rationality." (Lewis 1969: 55) In other words, provided it is not too hard to see that we have reason to believe that φ we will believe φ. Hence, on Cubitt and Sugden's account, we believe all tautologies that aren't too difficult to work out, and it may be doubted whether this prediction is plausible.

3.2. Sillari

Whereas Cubitt and Sugden's account is syntactic, Sillari (2005) adopts a semantic approach to prove Lewis's theorem. In this section, I will successively discuss how Sillari interprets the notions 'reason to believe' and 'indication', how he defines 'common reason to believe', and how he proves Lewis's theorem.

For Sillari 'i has reason to believe ___' is a modal operator R_i on propositions that has the following axiomatisation:[5]

$$(R_i\varphi \wedge R_i(\varphi \rightarrow \psi)) \rightarrow R_i\psi \tag{B1}$$

$$\text{From } \varphi \text{ infer } R_i\varphi \tag{B6}$$

(B1) is the K axiom of modal logic, which seems acceptable in this context. (B6) is the rule of necessitation that is standard in modal logic. It says that every agent has reason to believe all tautologies. Sillari argues that "it makes sense to require that an agent has reason to believe what logic dictates, although it is not the case that the agents in the model will come to *actually* believe all logical truths" (Sillari 2005: 289).

[5]Sillari also assumes that the operator has positive and negative introspection, which makes the accessibility relations transitive and euclidean. So he holds that the logic of having reasons to believe is K45.

https://doi.org/10.1017/S0266267123000238 Published online by Cambridge University Press

For reasons discussed at the end of the last subsection, this is potentially problematic.

The axiom (B1) and the rule (B6) make R_i a normal modal operator, for which Sillari gives a Kripke semantics. For each agent i there is a relation $\mathcal{R}_i$ on a set of possible worlds W. The semantics of $R_i\varphi$ is given in the usual way: $R_i\varphi$ is true in a world w in a model M if and only if φ is true in each world that is accessible from w:

$$(M, w) \models R_i\varphi \text{ iff } (M, v) \models \varphi \text{ for all } v \text{ such that } (w, v) \in \mathcal{R}_i. \tag{S1}$$

'Indication' is treated as a modal operator on pairs of propositions[6] with the following axiomatisation:[7]

$$(R_i\varphi \wedge (\varphi \rightarrow \psi)) \rightarrow Ind_i\, \varphi\psi \tag{B2}$$

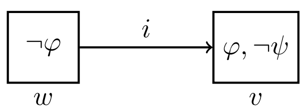
$$(R_i\varphi \wedge Ind_i\, \varphi\psi) \rightarrow R_i\psi \tag{B3}$$

(B3) is unproblematic: it is just the left-to-right part of Lewis's definition of indication. But (B2) may be too strong an assumption. It entails that agents have reasons to believe all logical consequences of everything they have reason to believe, which is not uncontroversial. Another objection to this axiom is that it does not fit with Lewis's definition of indication because that definition contains no implication from φ to ψ.

The semantics of $Ind_i\varphi\psi$ is given by[8]

$$(M, w) \models Ind_i\, \varphi\, \psi \text{ iff } (M, w) \models R_i\varphi \text{ and } (M, w) \models (\varphi \rightarrow \psi). \tag{S2}$$

This definition is inconsistent with (B3). To see this, note that while (B3) is equivalent to $Ind_i\, \varphi\, \psi \rightarrow (R_i\varphi \rightarrow R_i\psi)$, (S2) yields $Ind_i\, \varphi\, \psi \rightarrow (\varphi \rightarrow \psi)$. Now we construct a counterexample with two worlds, w and v:



According to (S2), we have $w \models Ind_i\, \varphi\, \psi$, because $w \models R_i\varphi$ and $w \models (\varphi \rightarrow \psi)$. So, according to (B3), we have $w \models R_i\psi$. On the other hand, according to (S1), $w \not\models R_i\psi$, because it is not the case that $v \models \psi$.[9]

This contradiction poses a problem because both (B3) and (S2) are essential to Sillari's account. The former represents Lewis's definition of indication. The latter plays an essential role in Sillari's semantical proof of Lewis's theorem. There is no obvious way, as far as I can see, to reformulate these principles in such a way that the problem is circumvented.

Clearly, this is a serious problem, but let us put it on one side and consider Sillari's proof of Lewis's theorem (Sillari 2005: 390-391). Sillari first introduces a new

---

[6]Sillari thus drops the distinction between 'states of affairs' and 'propositions'.

[7]Sillari uses the notation $R_i\varphi \Rightarrow_i R_i\psi$ but I will use the notation $Ind_i\varphi\psi$ instead that is employed everywhere else in this paper.

[8]Sillari uses $\neg(\varphi \wedge \neg\psi)$ instead of $\varphi \rightarrow \psi$, which is equivalent.

[9]Sillari adds the following note to (S2): "Notice that the clause for indication is restricted to the deductive aspect of indication only, since in this article non-deductive capabilities of the agents are not taken into account." This restriction is not relevant to the counterexample.

https://doi.org/10.1017/S0266267123000238 Published online by Cambridge University Press

modal operator CR_G for a group of agents G that stands for "members of G have common reason to believe that …". In order to conveniently express the semantic clause for CR_G, he uses the concept of G-reachability: "we say that a world v is G-reachable in k steps from world w iff there is a path of length k from v to w such that the edges between adjacent worlds are labelled by the accessibility relations of members of G". It follows that: "$(M, w) \models CR_G\,\varphi$ iff $(M, v) \models \varphi$ for all worlds v that are G-reachable from w in any numbers of steps."

The proof of Lewis's theorem proceeds by induction on the length of the path. Unfortunately, the inductive step of this proof is either incorrect or the formulation of the theorem is trivial, as is shown in detail in the appendix, and it is doubtful that this can be fixed.

In conclusion, Sillari's proposal is flawed in several respects. Most importantly, it uses two principles, (B3) and (S2), that are mutually inconsistent. Moreover, it uses assumptions that are too strong to be plausible and it replaces Lewis's argument with something entirely different, and therefore it hardly counts as an explication of Lewis's account.

In the next section, I will present a new reconstruction of Lewis's theory by developing the notion of 'reason to believe' a bit further than Lewis did.

4. Reasoning with reasons

Sillari and Cubitt and Sugden treat 'reason to believe' as an operator on propositions. This treatment misses an important aspect of Lewis's account, as the following quotes illustrate:

> … A indicates to i that j has reason to believe that ___ (*this reason* being *j's reason* to believe A holds) (Lewis 1969: 53, my emphasis).
>
> … A indicates to someone i that ___ if and only if, if i had reason to believe that A held, i would *thereby* have reason to believe that ___ (52-53, my emphasis).

In the first quote, Lewis identifies one reason with another, which suggests that reasons should be treated as objects, at least in the sense that two reasons can be the same or different. This suggestion is strengthened by the use of 'thereby' in the second quote, which seems to be saying that the reason to believe ___ is based upon the reason to believe that A holds.

A reason to believe something may be seen as a piece of evidence that supports that belief (Égré *et al.* 2021: 689). For example, seeing a bright flash of light followed by a loud bang is a reason to believe that a thunderstorm is underway. Or, having seen many people driving on the right and never one on the left, is a reason to believe that almost everyone drives on the right (cf. Lewis 1969: 40).

Based on these observations, I propose to analyse Lewis's argument using a logic that allows us to talk about reasons. My logic is a minimal extension of a subset of justification logic (Artemov and Fitting 2019; Kuznets and Studer 2019). Obviously, justifications and reasons are related concepts (Brandom 2009: 5), and therefore it shouldn't come as a surprise that a version of justification logic can be used to reason about reasons to believe.

It bears emphasising that justification logic was originally developed for quite different purposes than the one pursued here. In mathematical logic, it provides a semantics for intuitionistic logic in which justification terms represent proofs. In this context, $t : X$ represents that t is a proof of proposition X. In formal epistemology, justifications have been used to complement standard possible-worlds analyses of knowledge and belief. So, unlike the proposals discussed in the foregoing, the present account of reasons to believe is motivated entirely on independent grounds.

The language of our logic is based on a countable set of atomic propositions, a finite set of (names for) agents, and countable sets of constants and variables that represent reasons. The formulas of the language are defined inductively as follows:

1. each atomic proposition is a formula
2. if α and β are formulas then $\alpha \to \beta$ and $\alpha \wedge \beta$ are formulas
3. if α is a formula, i is an agent and r is a reason, then $r :_i \alpha$ is a formula. This formula is to be read as 'r is a reason for agent i to believe α'
4. if α is a formula and i is an agent, then $\exists r(r :_i \alpha)$ is a formula.

The logic features an operator, '$\cdot$', for constructing composite reasons:[10] if r and s are reasons, then $r \cdot s$ is also a reason. Composite reasons are used to define a version of modus ponens:

$$s :_i (\alpha \to \beta) \to ((t :_i \alpha) \to (s \cdot t :_i \beta))$$

If an agent i has reason s to believe the implication $\alpha \to \beta$ and reason t to believe α, then $s \cdot t$ represents his reason to believe β. So two reasons are needed to infer β. The operator is called 'application' because a reason to believe an implication may be seen as a device that applies to any reason to believe α so as to produce a reason to believe β.

Besides a set of axioms for classical propositional logic and the rule of modus ponens, our logic has the following application rule:[11]

$$s :_i (\alpha \to \beta),\ t :_i \alpha \ \vdash \ s \cdot t :_i \beta \tag{AR}$$

In contrast to the accounts of Cubitt and Sugden and Sillari, this logic does not require that agents have a reason to believe for every tautology. This premise is not needed to prove Lewis's theorem, and as we will see, a relatively weak logic suffices for a proof of that theorem: we only have to assume that agents have a reason to believe that $\alpha \wedge \beta$ follows from α and β, and a reason to believe that $\alpha \to \gamma$ follows

[10]Justification logic has a second operator '+' to form composite justifications with an associated deduction rule, which is not needed here. Besides, this rule would make the logic monotonic. That is undesirable because, contrary to proofs in mathematics, reasons to believe are defeasible. For applications of justification logics with this second operator to (justified) common knowledge, see Artemov (2006) and Bucheli *et al.* (2011).

[11]We formulate this principle as a rule and not as an axiom because we will use natural deduction proofs instead of axiomatic proofs.

https://doi.org/10.1017/S0266267123000238 Published online by Cambridge University Press

from $\alpha \rightarrow \beta$ and $\beta \rightarrow \gamma$. The appendix shows that these two assumptions are sufficient.

With this logic in place, we can give a new interpretation of the two basic concepts of Lewis's account.

To begin with, I will interpret 'i has reason to believe that φ' as 'i has *a* reason to believe that φ'. I will keep the notation $R_i\varphi$, but with a new meaning, which will help to compare our account with those of Cubitt and Sugden (2003) and Sillari (2005). The new meaning of $R_i\varphi$ is: 'there is *a* reason r such that $r :_i \varphi$'. So $R_i\varphi$ is an existential claim: that i has *a* reason to believe φ. Hence, whereas from $r :_i \varphi$ we can infer $R_i\varphi$, it is not possible to conclude $r :_i \varphi$ from $R_i\varphi$.

Let us now turn to the notion of indication. According to Lewis, $Ind_i\, A\varphi$ means that if i had a reason to believe that A holds, he thereby would have a reason to believe φ. Now, suppose i has a reason, say r, to believe that A holds and *thereby* has reason to believe φ. I interpret 'thereby' as meaning that the reason to believe φ is, at least partly, provided by r. In our logic, we can formalise this by saying that there is a reason s that, together with r, is the reason to believe φ. An obvious choice for the latter reason then is $s \cdot r$. Consequently, s is a reason to believe $A \rightarrow \varphi$.[12]

Therefore, we interpret 'the state of affairs A indicates to i that φ' as 'i has a reason to believe that A implies φ'. Accordingly, $Ind_i\, A\varphi$ is an abbreviation of 'there is a reason r such that $r :_i (A \rightarrow \varphi)$'. This interpretation closely follows what Lewis says about indication. In particular, this interpretation takes into account the word 'thereby' in Lewis's definition of indication quoted on page 4.

To demonstrate this, let us assume that A indicates φ to i. In our interpretation, this means that there exists a reason s such that $s :_i (A \rightarrow \varphi)$. Then we have to show that if i has a reason to believe that A holds he thereby has a reason to believe φ. Suppose i has a reason, say r, to believe that A holds. By rule (AR), we know that there indeed is a reason to believe φ, namely $s \cdot r$. This reason contains r as a constituent, which does justice to the suggestion of the word 'thereby' that the reason to believe φ is partly provided by the reason to believe that A holds.

Interestingly, this interpretation of indication is consistent with "in the logic of reasoning that i endorses, there is an inference rule which legitimates inferring x from 'A holds' " in Cubitt and Sugden (2003: 187).

Our interpretations of the concepts of reason to believe and indication enable us to close the gaps in Lewis's argument and to formally derive Lewis's theorem. We begin with reformulating the conditions for a state of affairs A that holds to be a basis for common knowledge of a proposition φ.

$$R_iA \rightsquigarrow \exists r(r :_i A) \tag{C1}$$

$$Ind_i\, A(R_j\, A) \rightsquigarrow R_i(A \rightarrow R_j\, A) \rightsquigarrow \exists r\big(r :_i \big(A \rightarrow \exists s(s :_j A)\big)\big) \tag{C2}$$

$$Ind_i\, A\varphi \rightsquigarrow R_i(A \rightarrow \varphi) \rightsquigarrow \exists r(r :_i (A \rightarrow \varphi)) \tag{C3}$$

[12]To be able to prove this in a correct formal way, we should replace the $\cdot$ operator in our logic with an annotated $\cdot_\varphi$ operator, as Renne (2009) and Kuznets and Studer (2013) do. Such an annotated operator makes it possible to formulate a dynamic logic in which reasons can also be eliminated. For this paper, there is no such added value and its use would unnecessarily complicate the discussion.

Lewis assumed that the agents have a reason to believe that they share their inductive standards and background information and therefore have a reason to believe that A indicates the same things to them. Cubitt and Sugden translated this assumption into their axiom (C4). By replacing all occurrences of $Ind_i\, A\psi$ in this axiom with $R_i(A \to \psi)$ we get:

$$\forall i, j \in P : R_i(A \to \psi) \to R_iR_j\ (A \to \psi) \tag{C4}$$

We will only need this assumption in so far as ψ is one of the propositions φ, $R_i\varphi$, $R_iR_j\varphi$, $R_iR_jR_k\varphi$, and so on.

Before spelling out Lewis's argument, we formulate three theorems of our logic that will simplify the proof:

$$R_i(\alpha \to \beta),\ R_i\alpha\ \vdash\ R_i\beta \tag{E1}$$

$$R_i(\alpha \to \beta),\ R_i(\beta \to \gamma)\ \vdash\ R_i(\alpha \to \gamma) \tag{E2}$$

$$R_iR_j\ (\alpha \to \beta)\ \vdash\ R_i\big(R_j\ \alpha \to R_j\ \beta\big) \tag{E3}$$

The proofs of these theorems, which can be found in the appendix, require only a relatively weak logic. In particular, there is no need to assume that the agents have a reason to believe for every tautology, as in the accounts of Cubitt and Sugden and Sillari.

The axiom schemas (A1) and (A6) that Cubitt and Sugden needed to prove Lewis's theorem can be easily deduced in our theory. The proofs can be found in the appendix.

Now, we can complete the proof of Lewis's theorem:

	— first-order reasons to believe	
(1)	$\forall i \in P : R_i\ A$	C1
(2)	$\forall i, j \in P : R_i\big(A \to R_jA\big)$	C2
(3)	$\forall i \in P : R_i(A \to \varphi)$	C3
(4)	$\forall i \in P : R_i\varphi$	3, 1, E1
	— second-order reasons to believe	
(5)	$\forall i, j \in P : R_iR_j(A \to \varphi)$	3, C4
(6)	$\forall i, j \in P : R_i\big(R_jA \to R_j\varphi\big)$	5, E3
(7)	$\forall i, j \in P : R_i\big(A \to R_j\varphi\big)$	2, 6, E2
(8)	$\forall i, j \in P : R_iR_j\varphi$	7, 1, E1
	— third-order reasons to believe	
(9)	$\forall i, j, k \in P : R_i\big(R_j(A \to R_k\varphi)\big)$	7, C4
(10)	$\forall i, j, k \in P : R_i\big(R_jA \to R_jR_k\varphi\big)$	9, E3
(11)	$\forall i, j, k \in P : R_i\big(A \to R_jR_k\varphi\big)$	2, 10, E2
(12)	$\forall i, j, k \in P : R_iR_jR_k\varphi$	11, 1, E1
	— higher-order reasons to believe	
(13)	etc.	

The last column of this proof shows the role of the different axioms and theorems. Axiom (C4) enables the step to the next higher-order proposition, while theorems (E2) and (E3) allow for this proposition to be transformed into a higher-order reason to believe. These theorems thus do the work that axiom (A6) did in Cubitt and Sugden's formalisation. At first sight, explicit reasons do not seem to play a role in this proof. However, their role is visible in the proofs of theorems (E2) and (E3) which can be found in the appendix.

This proof, like Lewis's own proof, provides just the first steps of an infinite series of steps. However, that is not sufficient for a proper proof. A rigorous proof by mathematical induction is given in the appendix.

In conclusion, I have given new interpretations of 'i has reason to believe' as 'i has *a* reason to believe' and of 'A indicates φ to i' as 'i has *a* reason to believe that A implies φ'. A major advantage of this approach is that the two basic concepts of 'having reason to believe' and 'indication' are reduced to a single concept 'reason to believe'. But more importantly, a minimal logic for reasoning about these reasons with minimal assumptions about inferences enabled us to close the gaps in the proof of Lewis's theorem.

5. From 'reason to believe' to 'actual belief'

A wise man proportions his belief to the evidence. (Hume 2007)

Lewis's theory aims to describe how we coordinate our activities and, therefore, it requires an understanding of beliefs. Specifically, the theory must explain how individuals convert their reasons to believe into actual beliefs because it is only actual beliefs that can shape their behaviour. According to Lewis, only the first few orders of actual belief will be formed because common belief, which is an infinite series of ever-higher orders of belief, is unattainable. This is because the iteration of reasons to believe is "a chain of implications, not of steps in anyone's actual reasoning" but actual reasoning is required to form actual beliefs, necessitating ancillary premises about rationality. In this context, rationality does not refer to decision-theoretic or game-theoretic concepts, such as maximizing expected utility or having unlimited deductive power. Instead, it signifies that someone who has reason to believe something will come to believe it. This kind of rationality comes in degrees, and the higher the order of a reason to believe, the higher the degree of rationality it takes to come to actual belief.

In this paper, three reconstructions of Lewis's theory of reasons to believe have been discussed. In this section, I will consider whether these reconstructions can shed some light on the transition of a reason to believe to an actual belief.

Cubitt and Sugden (2003) investigate what Lewis's analysis would look like if they added some strong assumptions about the agents' rationality, acknowledging that these assumptions are unrealistic and not in line with Lewis's intentions. The analysis does, however, lead to an interesting conclusion, namely that a basis A for common knowledge of a proposition φ may produce not only common reason to believe φ but also common (actual) belief of this proposition. This requires assumptions about the rationality of the agents, the basis for common knowledge, and the indication relation. First, every agent has to be a faultless reasoner, that is, he must believe everything he has reason to believe. Furthermore, this basis A must also be a basis for common knowledge that everyone in P reasons faultlessly. Finally, there are two ancillary premises constraining indication:

$$Ind_i\ A\varphi \wedge Ind_i\ A\psi \rightarrow Ind_i\ A(\varphi \wedge \psi) \tag{A3}$$

$$Ind_i\ A\varphi \wedge (\varphi \text{ entails } \psi) \rightarrow Ind_i\ A\psi \tag{A5}$$

With these assumptions in place, the agents will have common belief of φ.

Sillari (2005) models the difference between reasons to believe and actual beliefs by introducing an 'awareness structure' on the set of possible worlds. For each agent i there is a set A_i of formulas the agent is aware of. A formula $A_i\varphi$, i.e. agent i is aware of φ, means that agent i is sufficiently rational to believe φ if he has reason to believe φ. So an agent actually believes a formula if and only if two conditions are met: he has reason to believe it, and the formula belongs to his awareness set. This can be expressed as $B_i\varphi \equiv R_i\varphi \wedge A_i\varphi$.

The concept of degrees of rationality, as proposed by Lewis, can be modeled using the awareness structure. First-order rationality of an agent i then means that φ is in his awareness set. Second-order rationality means that all formulas of the form $B_j\varphi$ are in his awareness set. Now suppose that i also has reason to believe that j has first-order rationality. Then he will have the second-order beliefs $B_iB_j\varphi$. Similarly, it is possible to model any level of belief. While this account ties in with Lewis's explanation of the conversion of reasons to believe into actual beliefs, it lacks explanatory power because it reduces rationality to a list of formulas that are taken to be given (Paternotte 2011: 264).

These two accounts show that it is difficult to formalise Lewis's account of acquiring higher-order actual beliefs. Lewis's explanation starts with the generation of an infinite series of ever-higher orders of reasons to believe. He then describes the conversion of reasons to believe into actual beliefs. This conversion follows the structure of the series of reasons to believe. For each step in the conversion, an agent needs a higher degree of rationality and reasons to ascribe a higher degree of rationality to the other agents. Lewis does not spell out what kind of rationality is involved in this process. Neither of the two accounts succeeds in clarifying this account.

The reconstruction proposed in section 4 provides insight into why each step in the hierarchy of higher-order reasons to believe is an implication that does not require actual reasoning, whereas each step in the transition from reason to believe into actual belief requires an ever higher degree of rationality. To proceed from n-order reasons to believe to $n+1$-order reasons to believe only some weak principles of logic are needed. So, logically speaking, this transition is not complex at all. However, the reasons themselves become increasingly complex at each step. As can be seen in the proof of Lewis's theorem, each step requires the use of theorems (E2) and (E3), each of which uses the application rule (AR) that produces composite reasons. This might explain why the demands on the rationality required for the conversion from reasons to believe into actual beliefs increase rapidly. So, further research along this formalisation of Lewis's theory, might be useful.

Acknowledgements. For helpful discussions and critical comments on earlier versions of this paper, I would like to thank Bart Geurts, Tonny Hurkens, Jaap van der Does, and Robert Sugden. I am very grateful for the suggestions of an anonymous reviewer, who raised important objections and helped me clarify the paper.

Competing interests. None.

References

Artemov S. 2006. Justified common knowledge. *Theoretical Computer Science* **357**, 4–22. doi: 10.1016/j.tcs.2006.03.009.

Artemov S. and M. Fitting 2019. *Justification Logic - Reasoning with Reasons*. Cambridge: Cambridge University Press.

Brandom R. 2009. *Reason in Philosophy: Animating Ideas*. Cambridge, MA: Belknap Press of Harvard University Press.

Bucheli S., R. Kuznets and T. Studer 2011. Justifications for common knowledge. *Journal of Applied Non-Classical Logics* **21**, 35–60. doi: 10. 3166/jancl.21.35-60.

Clark H.H. and C.R. Marshall 1981. Definite reference and mutual knowledge. In *Elements of Discourse Understanding*, ed. A. Joshi, B. Webber and I. Sag, 10–63. Cambridge: Cambridge University Press.

Cubitt R.P. and R. Sugden 2003. Common knowledge, salience and convention: a reconstruction of David Lewis' game theory. *Economics and Philosophy* **19**, 175–210. doi: 10.1017/S0266267103001123.

Cubitt R.P. and R. Sugden 2014. Common reasoning in games - a Lewisian analysis of common knowledge of rationality. *Economics and Philosophy* **30**, 285–329. doi: 10.1017/S0266267114000339.

Égré P., P. Marty and B. Renne 2021. Knowledge, justification, and adequate reasons. *Review of Symbolic Logic* **14**, 687–727. doi: 10.1017/S1755020319000522.

Fagin R., J.Y. Halpern, Y. Moses and M.Y. Vardi 1995. *Reasoning about Knowledge*. Cambridge, MA: MIT Press.

Heifetz A. 1999. Iterative and fixed point common belief. *Journal of Philosophical Logic* **28**, 61–79. doi: 10.1023/A:1004357300525.

Hoek W. van der and J.J. Meyer 1997. A complete epistemic logic for multiple agents. In *Epistemic Logic and the Theory of Games and Decisions*, ed. M. Bacharach, L.-A. Gerard-Varet, P. Mongin and H. S. Shin, 35–68. Boston, MA: Springer.

Horty J.F. 2012. *Reasons as Defaults*. Oxford: Oxford University Press.

Hume D. 2007. *An Enquiry concerning Human Understanding*. Ed. P. F. Millican. Oxford: Oxford University Press.

Johnson-Laird P., R. Byrne and W. Schaeken 1992. Propositional reasoning by model. *Psychological Review* **99**, 418–439.

Kuznets R. and T. Studer 2013. Update as evidence: belief expansion. In *Proceedings of Logical Foundations of Computer Science LFCS'13, Lecture Notes in Computer Science, vol. 7734*, ed. S.N. Artemov and A. Nerode, 266–279. Berlin: Springer. doi: 10. 1007/978-3-642-35722-0_19.

Kuznets R. and T. Studer 2019. *Logics of Proofs and Justifications*. College Publications.

Lewis D. 1969. *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

Lewis D. 1978. Truth in fiction. *American Philosophical Quarterly* **15**, 37–46.

Moshman D. 2020. *Reasoning, Argumentation, and Deliberative Democracy*. New York, NY: Routledge.

Paternotte C. 2011. Being realistic about common knowledge: a Lewisian approach. *Synthese* **183**, 249–276. doi: 10.1007/s11229-010-9770-y.

Paternotte C. 2017. The fragility of common knowledge. *Erkenntnis* **82**, 451–472. doi: 10.1007/s10670-016-9828-4.

Renne B. 2009. Evidence elimination in multi-agent justification logic. In *Proceedings of the 11th Conference on Theoretical Aspects of Rationality and Knowledge – TARK '09*, 227. California: ACM Press. doi: 10.1145/1562814.1562845.

Schiffer S.R. 1972. *Meaning*. Oxford: Clarendon Press.

Sillari G. 2005. A logical framework for convention. *Synthese* **147**, 379–400. doi: 10.1007/s11229-005-1352-z.

Sperber D. and D. Wilson 1995. *Relevance: Communication and Cognition*. Second edition. Oxford: Blackwell.

Stalnaker R. 2014. *Context.* Oxford: Oxford University Press.
van Ditmarsch H., W. van der Hoek and B. Kooi 2008. *Dynamic Epistemic Logic.* Dordrecht: Springer.
Vanderschraaf P. 1998. Knowledge, equilibrium and convention. *Erkenntnis* **49**, 337–369. doi: 10.1023/A:1005461514200.

Appendix

This appendix contains the proofs of the theorems E1, E2, E3, A1, A6, and Lewis's theorem and the refutation of Sillari's proof of Lewis's theorem. The proofs of E1 - E3 are Fitch-style natural deduction proofs. They use the usual rules for natural deduction proofs for propositional logic and the introduction and elimination rules for the existential quantifier. In order to make this section self-contained, I present the latter two rules.

Introduction rule for the existential quantifier

$$\begin{array}{rll} m. & [\tau/\alpha]\varphi & \\ n. & \exists\alpha\,\varphi & \mathrm{I}\exists\ m \end{array}$$

Elimination rule for the existential quantifier

$$\begin{array}{rll} m. & \exists\alpha\,\varphi & \\ n. & \quad [\gamma/\alpha]\varphi & \text{ass} \\ o. & \quad \chi & \\ o+1. & \chi & \mathrm{E}\exists\ m \end{array}$$

In this rule, γ is a constant which does not occur in φ or χ, or in any assumption which is undischarged in the derivations up to line m (other than the assumption in line n).

Proof of theorem E1

Theorem E1 is a direct consequence of the application rule.

Theorem 1 (E1). $R_i(\alpha \to \beta),\ R_i\alpha \vdash R_i\beta$

Proof.

$$\begin{array}{rll} 1. & R_i(\alpha \to \beta) & \text{ass} \\ 2. & R_i\alpha & \text{ass} \\ 3. & \quad s :_i (\alpha \to \beta) & \text{ass} \\ 4. & \quad\quad t :_i \alpha & \text{ass} \\ 5. & \quad\quad s \cdot t :_i \beta & \text{AR } 3,\ 4 \\ 6. & \quad\quad R_i\beta & \mathrm{I}\exists\ 5 \\ 7. & \quad R_i\beta & \mathrm{E}\exists\ 2 \\ 8. & R_i\beta & \mathrm{E}\exists\ 1 \end{array}$$

□

Proofs of theorems E2 and E3

The proofs of the theorems E2 and E3 don't require very strong assumptions. In particular, it is not necessary to assume that the agents have reasons to believe every propositional tautology. The following two assumptions about reasons to believe propositional tautologies will be needed:

$$\exists a(a :_i (\alpha \to (\beta \to (\alpha \wedge \beta)))) \tag{T1}$$

$$\exists b(b :_i (((\alpha \to \beta) \wedge (\beta \to \gamma)) \to (\alpha \to \gamma))) \tag{T2}$$

Note that we do not assume that every agent has reason to believe every tautology. The only two tautologies for which it is required that every agent has a reason to believe are two quite basic ones.[13]

In addition to these two assumptions, we need to assume that the agents have a reason to believe theorem (E1), the direct consequence of the application rule.[14]

$$\exists c\big(c :_i \big(R_j\ (\alpha \to \beta) \to \big(R_j\ \alpha \to R_j\ \beta\big)\big)\big) \tag{T3}$$

The logic we use is relatively weak, having characteristics that distinguish it from stronger logics. First, it is only a part of justification logic because it lacks the + operator and the corresponding axioms. Second, it avoids logical omniscience by only requiring reasons to believe a limited number of tautologies. Third, it is non-monotonic, meaning that it allows for reasons to believe two contradictory propositions.

Lemma (L1). $R_i\alpha,\ R_i\beta\ \vdash\ R_i(\alpha \wedge \beta)$

Proof. Now the assumptions $R_i\alpha$ and $R_i\beta$ are the existential formulas $\exists r(r :_i \alpha)$ respectively $\exists r(r :_i \beta)$, so we start the deduction by introducing a constant s and the assumption that it is a reason for i to believe α and a constant t and the assumption that it is a reason for i to believe β.

Then we try to introduce the conclusion, also an existential formula, by finding a term τ and showing that it is a reason for i to believe $\alpha \wedge \beta$:

$$
\begin{array}{rll}
1. & R_i\alpha & \text{ass}\\
2. & R_i\beta & \text{ass}\\
3. & \mid\ s :_i \alpha & \text{ass}\\
4. & \mid\ \mid\ t :_i \beta & \text{ass}\\
5. & \mid\ \mid\ a :_i (\alpha \to (\beta \to (\alpha \wedge \beta))) & \text{T1}\\
6. & \mid\ \mid\ a \cdot s :_i (\beta \to (\alpha \wedge \beta)) & \text{AR 5, 3}\\
7. & \mid\ \mid\ a \cdot s \cdot t :_i (\alpha \wedge \beta) & \text{AR 6, 4}\\
8. & \mid\ \mid\ R_i(\alpha \wedge \beta) & \text{I}\exists\ 7\\
9. & \mid\ R_i(\alpha \wedge \beta) & \text{E}\exists\ 2\\
10. & R_i(\alpha \wedge \beta) & \text{E}\exists\ 1
\end{array}
$$

□

Theorem 2 (E2). $R_i(\alpha \to \beta),\ R_i(\beta \to \gamma)\ \vdash\ R_i(\alpha \to \gamma)$.

Proof.

$$
\begin{array}{rll}
1. & R_i(\alpha \to \beta) & \text{ass}\\
2. & R_i(\beta \to \gamma) & \text{ass}\\
3. & R_i((\alpha \to \beta) \wedge (\beta \to \gamma)) & \text{L1 1, 2}\\
4. & \mid\ s :_i ((\alpha \to \beta) \wedge (\beta \to \gamma)) & \text{ass}\\
5. & \mid\ b :_i (((\alpha \to \beta) \wedge (\beta \to \gamma)) \to (\alpha \to \gamma)) & \text{T2}\\
6. & \mid\ b \cdot s :_i (\alpha \to \gamma) & \text{AR 5, 4}\\
7. & \mid\ R_i(\alpha \to \gamma) & \text{I}\exists\ 6\\
8. & R_i(\alpha \to \gamma) & \text{E}\exists 3
\end{array}
$$

□

[13]Johnson-Laird *et al.* (1992: 428) present evidence that reasoning with conjunction is easier than reasoning with any other connective. Moshman (2020: 23) presents evidence that preschool children can already draw logical inferences concerning transitivity.

[14]Taken together, the axioms (T1)–(T3) are consistent because having reasons to believe tautologies cannot create a contradiction.

Theorem 3 (E3). $R_iR_j\,(\alpha \to \beta) \vdash R_i\big(R_j\,\alpha \to R_j\,\beta\big)$.

Proof.

1.	$R_iR_j\,(\alpha \to \beta)$	ass
2.	$s :_i R_j\,(\alpha \to \beta)$	ass
3.	$c :_i (R_j\,(\alpha \to \beta) \to (R_j\,\alpha \to R_j\,\beta))$	T3
4.	$c \cdot s :_i (R_j\,\alpha \to R_j\,\beta)$	AR 3, 2
5.	$R_i(R_j\,\alpha \to R_j\,\beta)$	I∃4
6.	$R_i(R_j\,\alpha \to R_j\,\beta)$	E∃1

Proof of theorem A1

Theorem 4 (A1). $(Ind_i\;A\;\varphi \wedge R_i\;A) \to R_i\;\varphi$

Proof.

1.	$Ind_i\,A\,\varphi$	ass
2.	$R_i\,A$	ass
3.	$R_i(A \to \varphi)$	1, definition of indication
4.	$R_i\,\varphi$	3, 2, E1

□

Proof of theorem A6

Theorem 5 (A6). $\big(Ind_i\;A\big(R_jA\big) \wedge R_i\big(Ind_jA\varphi\big)\big) \to Ind_i\;A\big(R_j\;\varphi\big)$

Proof.

1.	$Ind_i\,A(R_jA)$	ass
2.	$R_i(Ind_jA\varphi)$	ass
3.	$R_i(A \to R_jA)$	1, definition of indication
4.	$R_iR_j(A \to \varphi)$	2, definition of indication
5.	$R_i(R_jA \to R_j\varphi)$	4, E3
6.	$R_i(A \to Rj\varphi)$	3, 5, E2
7.	$Ind_i\,A(R_j\,\varphi)$	6, definition of indication

□

Proof of Lewis's theorem

Lewis's theorem states: if a state of affairs A is a basis for common knowledge of a proposition φ in a group P (that is, it satisfies the conditions C1, C2 and C3), then there is common knowledge of φ, i.e. all finitely nested formulas $R_iR_j \ldots R_k\varphi$ hold for all $i, j, \ldots, k$ in P.

In order to provide a complete proof of this theorem, we need two definitions and a lemma. The first definition is a recursive definition that captures the set of formulas to which (C4) applies. The second definition formalises the expression 'all finitely nested formulas $R_iR_j \ldots .R_k\varphi$ for all agents $i, j, \ldots, k$ in P'.

Definition (Set of propositions Φ_1)

(1) $\varphi \in \Phi_1$
(2) $\alpha \in \Phi_1 \to \forall i \in P : R_i\alpha \in \Phi_1$

https://doi.org/10.1017/S0266267123000238 Published online by Cambridge University Press

Definition (Set of propositions Φ_2)

$$\Phi_2 = \Phi_1 \backslash \{\varphi\}$$

Lemma (L2). $\forall i \in P, \psi \in \Phi_1 : R_i(A \rightarrow \psi)$

Proof. We will use a proof by induction.

We first prove the base case, that is the statement $\forall i \in P : R_i(A \rightarrow \varphi)$. This statement is true because of (C3).

For the inductive step, we prove that, if $R_i(A \rightarrow \psi)$ is true, then $\forall j \in P : R_i(A \rightarrow R_j\psi)$ is true.

So, assume $i \in P$ and $R_i(A \rightarrow \psi)$.

Then, according to (C4), for every $j \in P$ we have $R_iR_j(A \rightarrow \psi)$.

So, according to (E3), we have $R_i(R_jA \rightarrow R_j\psi)$.

By (C2) and (E2), it follows that $R_i(A \rightarrow R_j\psi)$.

This completes the proof of the lemma. □

Theorem 6 (Lewis's theorem). *Every proposition in* Φ_2 *is true.*

Proof. Let $\alpha \in \Phi_2$. Then there is a $\beta \in \Phi_1$ and an $i \in P$ such that $\alpha = R_i\beta$. Then, according to the lemma, we have $R_i(A \rightarrow \beta)$. From (C1) and (A1), it follows that $R_i\beta$ is true, hence that α is true. This completes the proof. □

Evaluation of Sillari's proof

Proposition 4.1 (Sillari 2005: 391) says that, if the following conditions hold:

(a) $\forall i \in G : R_i\, \varphi$
(b) $\forall i, j \in G : Ind_i\, \varphi\, (R_j\, \varphi)$
(c) $\forall i \in G : Ind_i\, \varphi\, \psi$

then the agents in a group G have common reason to believe ψ.

Sillari's proof runs as follows.

> We show by induction on the length of the path that if v is G-reachable from the actual world w, then $(M, v) \models \psi$. Let v be G-reachable from w in 1 step. By B3 at w all of $R_G\varphi$, $R_GR_G\varphi$, and $R_G\psi$ hold. Hence, at v all of φ, $R_G\varphi$, and, as desired, ψ hold. By the induction hypothesis, if u is G-reachable from w in n steps, then all of φ, $R_G\varphi$ and ψ hold at u. However, from (b), (c) and the fact that $R_G\varphi$ holds at u, it follows that all of $R_G\varphi$, $R_GR_G\varphi$, and $R_G\psi$ hold at u, or that ψ hold at every world $u + 1$ which is reachable in one step from u. (391)

The proposition, however, does not clearly state if the conditions (a), (b) and (c) hold at every possible world or only at the world, say w, in which the agents have common reason to believe ψ. The former seems to be the case because the proof uses (b) and (c) not only in w but also in the world u in the induction hypothesis. But that means the whole proof by induction is superfluous. For if the conditions hold at all worlds, we can easily prove that ψ is true in every world v that is reachable from w in a finite number of steps. Let u be a world from which v is accessible. In u we have $R_i\varphi$ and $Ind_i\varphi\psi$. Hence, from (B3), we have $R_i\psi$. From (S2) now follows that ψ is true in v, which completes the proof that the agents have common reason to believe ψ.

In short, either the proof by induction is not correct, or the proposition is trivial.

Huub Vromen is a lecturer in logic at the Faculty of Philosophy, Theology and Religious Studies of Radboud University. He holds degrees in Mathematics and in Philosophy at Radboud University, and degrees in Law and in Management Sciences at the Open University. His research interests include the interface of logic and philosophy of language, and the use of proof assistants in philosophy.

Cite this article: Vromen H. Reasoning with reasons: Lewis on common knowledge. *Economics and Philosophy*. https://doi.org/10.1017/S0266267123000238