Formal Ontology in Information Systems
B. Brodaric and F. Neuhaus (Eds.)

doi:10.3233/FAIA200662

A Commonsense Theory of Secrets

Haythem O. ISMAIL [a] and Merna SHAFIE [b]
[a] *Cairo University and the German University in Cairo, Egypt, haythem.ismail@guc.edu.eg*
[b] *The German University in Cairo, Egypt, merna.shafiee@guc.edu.eg*

Abstract. With the advent of social robots, precise accounts of an increasing number of social phenomena are called for. Although the phenomenon of secrets is an important part of everyday social situations, logical accounts of it can only be found, in a rather strict sense, within logical investigations of systems security. This paper is an attempt to formalize the logic of a commonsense notion of secrets as a contribution to ontologies of social and epistemological phenomena. We take a secret to be a five-way relation between a proposition, a group of secret-keepers, a group of nescients, a condition of secrecy, and a time point. A bare-bones notion of secrets is defined by providing necessary and sufficient conditions for said relation to hold. Special classes of secrets are then identified by considering an assortment of extra conditions. The logical language employed formalizes a classical account of belief and intention, a theory of groups, and a novel notion of revealing. In such a rich theory, interesting properties of secrets are proved.

Keywords. Secrets, Commonsense reasoning, Belief, Intention, Revelation

1. Introduction

Secrets come in all shapes and sizes: They can be classified military maps, family-devastating incidences of spouse infidelity, critical credit card pin numbers, questions in an exam, names of academy award winners, locations of treasures, sorcerous procedures for invisibility, or embarrassing childhood mischiefs. Secrets are often hard to keep, yet they are sometimes gratifying to be part of. They are catalysts for suspicion, but they are also gauges of trust. Some secrets are a social necessity, and most are psychological burdens [1,2,3].

The list of intuitions about and curiosities of secrets can go on and on. But we are not concerned here with enumerating them, nor are we willing to analyze most of them. Our objective is more modest and more fundamental. For, while secrets are social/psychological phenomena par excellence, they have an obscure ontological/epistemological flip side. Guided by some foundational intuitions about secrets, we seek to arrive at a commonsense theory of secrets which is precise enough to be amenable to logical analysis. Such *logic of secrets* should be a necessary component of a logic-based artificial intelligence system which is expected to competently engage in social interaction with people. In the near future, social robots may be everywhere around us, assisting us at work, at hospitals, with house chores, and granted the status of trusted life partners [4]. These robotic companions should be capable of understanding what secrets are and of keeping our secrets.

As far as we know, studies of secrecy, within the logical tradition, are confined to issues related to system security [5,6,7,8,9,10,11,12,13,14]. In such studies, a secret is presented as a *true* piece of information about an agent/a system which is not known by

a certain adversary group; the focus is mostly on identifying effective, and often subtle, methods for keeping the secret. Several aspects of secrecy are not considered by these studies. For example, there is no account of a secret keeper's intention to keep the secret (which is *the* defining characteristic of secrecy [3]), no investigation of what it means for a secret to be revealed to someone (which does not, in general, effect knowledge or belief), and no discussion of the possible relations among keepers of the same secret.

The paper is structured as follows. Section 2 discusses ten commonsense intuitions about secrets and Section 3 motivates the novel notion of revelation. Section 4 presents a logical language for reasoning about secrets. Section 5 includes a number of theorems proved using the introduced logic. A typology of secrets is presented in Section 6. Finally, Section 7 reviews related work.

2. Ten Intuitions About Secrets

We would like to start our investigation by drawing a distinction between secrets on one hand and the *objects of* secrets on the other. That a particular military map is classified is not, in general, an intrinsic property of the map itself [15]; rather, it is an extrinsic property that the map acquires (possibly temporarily) by virtue of standing in a complex relation to other entities, notably a secret keeper or a group thereof. Further, it is possible that the same map is kept secret by General *A* from their spouse and, simultaneously and independently, by General *B* from their own spouse. We would like to say that there are *two* secrets here, both having the same map as their object, and that, for example, one secret may be divulged and the other not. While English locutions such as "This map is a secret" are common, a careful analysis of secrecy should not conflate secrets and their objects. In particular, the existence of objects of secrecy is a necessary but an insufficient condition for the existence of secrets. Henceforth, we shall refer to objects of secrecy as "secreta" (plural of "secretum") and shall take secrets to be relations between secreta and other entities—this is our first intuition.[1]

I1. Secrets are distinct from secreta; they are relations between secreta and other entities.

This, then, leads to the following question: What kind of entity can a secretum be? It would seem that all sorts of beasts can be secreta: pin numbers, names of academy award winners, military maps, recipes for invisibility, etc. We contend, however, that this display of diversity is an artifact of the elliptical language we use to talk about secrets. It is not, for example, the credit card pin number itself that is the secretum; the pin number may happen to be the date of birth of the card owner who can write it down on a sheet of paper, show it to everybody, and say that it is their birth date without thus revealing their secret pin number. Rather, the secretum is *the proposition that this number is the credit card pin number*. Similarly, prior to announcing the names of the academy award winners in 2019, a name such as "Rami Malek" was by no means a secretum, the secretum was *the proposition that"Rami Malek" is the name of the winner of the Best Actor award*. Henceforth, we uphold the following intuition.[2]

[1]This is similar to distinguishing the object of an intention–an action, for instance–from the intention itself which is a complex attitude an agent holds towards that object; likewise for belief and the object of belief.

[2]Some readers may be suspicious about **I2** due to examples such as the classified military map, where it makes sense for the responsible General *A* to physically hide the map itself from their spouse. While we agree that hiding the map is indeed the right thing to do, we do not agree that this makes the map itself a secretum.

I2. Secreta are propositions.

We now turn to the "entities", referred to in **I1**, whose standing in some relation to a secretum constitutes a secret. First, for a secret to exist, there must be some *secret keepers*. It makes little sense to claim that there is a secret which no one is keeping. Mere unawareness of a proposition or concealment thereof does not make it a secretum. For example, before discovering that the earth is spheroid, no one was aware of this fact. Nevertheless, it is hardly acceptable to claim that the earth's roundness was a secretum at that time, primarily because no one could *keep* it a secret since no one was aware of it in the first place. Similarly, that raw gold is *hidden* within some mountain is no secret until somebody discovers it and decides to keep it to themselves.

I3. For every secret, there is a group of secret keepers.

Not only do secrets require someone to keep them but they also require someone to be kept *from*. A person who is cast away on a deserted island with no hope of getting rescued cannot be said to be keeping any secrets simply because there is no one to hide it from. This is so even though whatever happens to them on the island is completely concealed from everyone. Hence, whenever there is a secret, there is a group from whom the secret is kept; we refer to its members as *nescients*.

I4. For every secret, there is a group of nescients.

Now, let us clarify what we mean by "group" in **I3** and **I4**. Each group is identified by a *group condition*. At any point in time, the set of group members is the set of agents satisfying the group condition at this time. An *extensional* group is one for which the group condition is membership in a certain *set*. Since members of a set are fixed over time, the members of an extensional group never change. An *intensional* group is a group which is not extensional. (Extensional groups are similar to the "plural individuals" of [16,17] or the "E-collectives" of [18]; intensional groups are the "groups" of [16,17] or the "I-collectives" of [18]). Secret keepers and nescients could be of either group type. For example, a crush on a high-school colleague is possibly a personal secret with an extensional (singleton) group of secret keepers. On the other hand, an esoteric sorcerous procedure for invisibility may be kept secret by an intensional group of sorcerers who keep on handing it down for centuries across generations. In both examples, nescients form the (extensional and, respectively, intensional) group which includes all those who are not secret keepers. Extensional and intensional groups may be empty. An extensional group is empty only if the corresponding set is empty. Such a (unique) group is *necessarily forever empty* (NFE). An intensional group may become temporarily empty for a certain period during which no one satisfies the group condition. If the group condition is such that, starting at some time, it is necessarily the case that no one would ever satisfy its condition, then this intensional group is also NFE. It is necessary for the existence of secrets that the group of keepers is not NFE, but it may (if intensional) be temporarily empty, and that the group of nescients is not believed to be NFE by any of the keepers.

Rather, the secretum is the proposition that the map is a map of some critical military site. The spouse's finding the map causes the revelation of *the proposition that there is this strange map* which, given the nature of the Genaral's work, may lead to the conclusion that it is a map of some military site.

I5. Secret keepers and nescients form extensional or intensional groups. The group of keepers is not necessarily forever empty and none of its members believe that the group of nescients is necessarily forever empty.

Secrets are, in general, not eternal [19]. Most secrets are only kept so long as some condition of secrecy holds. For example, as per the "Automatic Declassification Program" in the United States of America,

> Information appraised as having permanent historical value is automatically declassified once it reaches 25 years of age unless an agency head has determined that it falls within a narrow exemption that permits continued classification and it has been appropriately approved. [20]

I6. A proposition is only a secretum as long as some condition of secrecy holds.

The condition of secrecy is a condition on the persistence of the keepers' intention to keep the secret. It often happens, however, that a secret is (accidentally or ill-intentionally) exposed to a nescient when the condition of secrecy still holds. These are cases in which the secret keepers fail to keep the secret. Compare, for example, between the natural expiration of a secret exam, which happens when students sit for the exam, and its premature exposure as a result of a malicious student's gaining access to the professor's computer. Hence, secrets are temporary in the stronger sense that, regardless of the secrecy condition, they may fail to be kept as long as intended.[3]

I7. Secrets are temporary.

Hence, we take secrets to be four-way relations which temporarily hold between a secretum ϕ, a group K of secret keepers, a group N of nescients, and a condition C of secrecy. Formally, we write $Secret(\phi, K, N, C, t)$ to state that, at time t, proposition ϕ is kept a secret by the group K from the group N while proposition C holds. Exactly what conditions are necessary and sufficient for said relation to hold is what we now turn to.

First, in a genuine secrecy situation, all members of K believe ϕ [21] and C—otherwise they will have no motivation for keeping ϕ a secret.

I8. Secret keepers believe both the secretum and the secrecy condition.

A possible objection to **I8** is that people often confide their secrets to others [22]. For example, x may inform y about their secretum ϕ and ask them to never reveal it to anybody. While y may fail to believe ϕ, they, nevertheless, form the intention of never mentioning it to anyone. Can we then say that the (extensional) group formed of x and y is keeping ϕ a secret from everyone else? We do not think so. Note that y's intention to never mention ϕ to anyone may be based solely on their commitment to honesty; since y does not believe ϕ, it would be deceptive to state it [23]. Thus, the situation here is indistinguishable from one in which y is simply being honest; it would be awkward to claim that every time we decide to not tell a lie we are keeping a secret. Even if y is generally dishonest, and are more than willing to lie about ϕ, but they do not do it out of respect for x's wishes, it is

[3] We follow [19] in taking "temporary" to qualify a phenomenon as being not necessarily permanent. Hence, while secrets are, in general, temporary, some secrets may *happen* to be kept permenantly.

still not plausible to claim that they are keeping ϕ a secret. We have all sorts of reasons for not saying certain things (especially if we do not believe them); among other things, we do it to be respectful, polite, suspenseful, and even spiteful. It is hardly acceptable to claim that in all these situations we are keeping a secret.[4] Notwithstanding the above argument, we are not saying that y is not keeping any secrets here; y is indeed keeping a secret, the secretum is not ϕ, but that ϕ *is a secretum of* x, which y indeed believes.

The second ingredient of the secrecy relation is that none of the secret keepers believes that the secretum has been revealed to a nescient. If they do, they will have no reason to continue keeping the secret.

I9. No secret keeper believes that the secretum has been revealed to a nescient.

There are at least two things to say about **I9**. First, sometimes a single nescient n gets to know about ϕ. This often does not result in the secret keepers' publicly disclosing the secret; they may choose to continue keeping it from the rest of the nescients. This is, however, not a counterexample to **I9**, for the secret has undergone a major change following the revelation to n. In particular, thenceforth, the group of nescients has changed into a group which does not contain n—resulting in a new secrecy relation. Second, **I9** is necessary to rule out certain situations which would otherwise be, counter-intuitively, counted as secrets. For example, n's friends, aware of how much weight-conscious they are, may decide, out of sheer courtesy, to never point out n's recent, visible weight gain. This is not a case of secrecy exactly because everyone knows that n is aware of the gain in their weight. The final ingredient of secrecy, and the most fundamental [3], is the keepers' *intention* to indeed keep the secret.

I10. Every secret keeper intends that the secretum is not revealed to a nescient as long as the secrecy condition holds.

Independently-motivated properties of intention yield intuitive properties of secrets. For example, according to [24], one cannot intend a proposition if they believe it to be false. Thus, in normal circumstances, it would be futile to keep the name of the capital of Georgia a secret since anybody can easily gain access to this public piece of information.[5]

In Section 4, we present a logical language in which we formalize the definition of a secret with regards to the previously mentioned intuitions. First, however, we need to elucidate the central notion of revelation.

3. Revelation

What does it mean for ϕ to be revealed to n? A prototypical revelation scenario is one in which an agent A truthfully states the true proposition ϕ to n, who does not know ϕ, thereby resulting in n's coming to believe ϕ. Not all instances of revelation, however, share the features of this idealized situation. First, whereas typical uses of the English "reveal" seem to indeed presuppose the truth of the revealed proposition [25], in our analysis of secrets (particularly, **I9** and **I10**), we do not want to assume that ϕ is true. Hence, the

[4] A parent, for example, may not want their young, gullible children to get exposed to some racist doctrines, which the parent does not believe in, lest they may subconsciously adopt them. We would not say that such doctrines are secrets *of* the parent but that they rather be kept unrevealed to their children.

[5] In *abnormal* circumstances, where the secret keeper can lock up the nescient and isolate them from the rest of the world, such a secret would be possible.

notion of revelation we need here does not carry this particular presupposition of the English verb. Second, we would like to capture a notion of revelation which does not assume that the revealer believes ϕ. Someone who is keeping ϕ a secret from n would take their secret to have been divulged following A's revelation, even if A does not believe ϕ and is attempting to mislead n. Third, ϕ's being revealed to n need not necessarily imply n's believing ϕ. For example, consider a professor who is keeping the questions of an exam secret from their students, but not from their assistant. Now suppose that the assistant discloses the contents of the exam to a student. The student, however, does not believe the assistant, thinking that they must be misleading them. In this situation, the professor would still consider their secret to have been divulged and would, typically, change the exam. Finally, in many cases, there is no agent A who reveals ϕ to n; mere perception of a state of affairs by n may be sufficient for the revelation of ϕ.

We are, thus, left with a very weak notion of revelation: "ϕ is revealed to n" means that n was somehow (possibly via perception) informed about ϕ. Revelation is not vacuous, though; it is strictly stronger than mere *awareness* [26]. For example, prior to announcing names of the academy award winners in 2019, everybody (who was interested) entertained, and was thus *aware*, of the proposition that Rami Malek is the winner of the best actor award; but this proposition was only *revealed* during the ceremonies. Thus, revelation is strictly stronger than awareness but strictly weaker than belief. We propose to intuitively construe ϕ's revelation to n as n's having (positive) evidence for ϕ. This being said, revelation is, thus, a special kind of modality. In particular, one can have evidence for both ϕ and $\neg\phi$; hence, both propositions may be revealed. We take this intuition up more seriously below by modeling revelation along the lines of the logic of evidence of [27].

4. Formalizing Secrets

To formalize secrets, we use a language $\mathscr{L}_S$ based on (a fragment of) the language VEL of [28], equipped with a special sort for groups, two normal modal operators for belief and intention, and a non-normal modal operator, akin to the evidence operator of [27], for revelation. Limitations of space allow us to only provide a sketch of the syntax and semantics of $\mathscr{L}_S$.

$\mathscr{L}_S$ is a sorted, first-order language with equality. In particular, there is a sort σ_A for agent-denoting terms, a sort σ_G for group-denoting terms, and a sort σ_T for time-denoting terms. A set of $\mathscr{L}_S$-atoms is generated in the usual way from countable sets of predicate symbols, function symbols, and variables. A special function symbol $[\cdot]$ combines with a term of sort σ_A to form a term of sort σ_G. Function symbols $\sqcup$ and $\sqcap$ form terms of sort σ_G from pairs of σ_G terms. Intuitively, $[A]$ denotes the extensional group comprised of the single member denoted by A, $G_1 \sqcup G_2$ and $G_1 \sqcap G_2$ denote the groups whose sets of members at any time are, respectively, the union and intersection of the sets of members of G_1 and G_2. A special binary predicate symbol Mem forms an atom by combining with terms of sorts σ_A and σ_G, respectively; intuitively, $Mem(A,G)$ means that agent A is a member of group G. Moreover, we have atoms of the form $\alpha = \beta$ (with the obvious semantics), where α and β are of the same sort, atoms of the form $t_1 \leq t_2$, where t_1 and t_2 are of sort σ_T, which mean that time point t_1 is no later than time point t_2, and atoms of the form $AT(t)$ which mean that the time (of evaluation) is t. $\mathscr{L}_S$ is the smallest set of formulas generated by the following grammar (and respecting the signatures of the predicate and function symbols).

$$\phi := P \mid \neg\phi \mid \phi \wedge \phi \mid \forall x[\phi] \mid \circledast\phi \mid H(\phi,t) \mid B(A,\phi) \mid I(A,\phi) \mid R(A,\phi)$$

where P is an atom, A is of sort σ_A, and t is of sort σ_T. Other logical connectives and the existential quantifier are defined in the standard way.

Expressions of $\mathscr{L}_S$ are interpreted over a branching tree structure. Each node in the tree is referred to as a *state*, and every state has a unique past and several possible futures [28]. A complete branch through the tree is a *history*, which is a bijection from a linearly-ordered set of time points to the set of states. Thus, one can view a history-time pair (h,τ) as a state. All expressions of the language are interpreted at such a pair (h,τ). In particular, where $\mathscr{V}$ is a valuation of the terms and the atoms, $H(\phi,t)$ means that "ϕ holds at t" and $[\![H(\phi,t)]\!]^{\mathscr{V}}_{h,\tau}$ is true if and only if $[\![\phi]\!]^{\mathscr{V}}_{h,[\![t]\!]^{\mathscr{V}}_{h,\tau}}$ is true. The expression $[\![\boxtimes\phi]\!]^{\mathscr{V}}_{h,\tau}$ is true if $[\![\phi]\!]^{\mathscr{V}}_{h',\tau}$ is true at all histories h' that coincide with h up to τ.

Formulas of the form $B(A,\phi)$ and $I(A,\phi)$, respectively, mean that "agent A believes ϕ" and "agent A intends ϕ", and are interpreted in the standard way using accessibility relations, one for each agent, on the set of history-time pairs (h,τ). A formula $R(A,\phi)$ intuitively means that ϕ is revealed to A. Following [27], we interpret revelation formulas using a function $\mathscr{R}$ which maps every agent and history-time pair (h,τ) to a family of sets of history-time pairs (h,τ) (each set, intuitively, corresponding to a proposition which is revealed to the agent in the history-time pair (h,τ)). Two important constraints on these families is that none of them is empty (tautologies are all revealed) or contains the empty set (contradictions are never revealed). Crucially, the families of sets are not closed under intersection, allowing agents to have contradictory propositions revealed to them without commitment to the revelation of falsehood.Thus, R is not a normal modal operator [29]. Hence, following [27], $[\![R(A,\phi)]\!]^{\mathscr{V}}_{h,\tau}$ is true if and only if there is some $X \in \mathscr{R}([\![A]\!]^{\mathscr{V}}_{h,\tau},h,\tau)$ such that $[\![\phi]\!]^{\mathscr{V}}_{h',\tau'}$ is true for every $(h',\tau') \in X$.

Note that the notion of revelation proposed here is a *passive* one; our modal operator R informally stands for what it means for a proposition to be revealed (in the sense of its being exposed or not covered) to an agent. We are not accounting for *acts* of revelation. As such, R is akin to B, and our account does not explain how revelation is caused just as no common account of belief investigates events that result in belief.

An axiomatic system, referred to as Σ, that captures the basic intuitions we have about the meaning of the various operators is displayed in Figure 1. (Variables are universally-quantified with widest scope unless otherwise indicated.)[6] Though not crucial for proving our theorems, we include (in the right column) axioms for the VEL [28] fragment we employ for completeness. As is common, B is a $KD45$ and I is a KD modal operator. **IB1** and **IB2** indicate that agents are never wrong about having or lacking intentions. **IB3** is motivated by [24]. It captures the intuition that intentions should be dropped once it is realized that they are impossible to achieve.

R1 indicates that tautologies are revealed and contradictions are not, **R2** requires revelation to be closed under logical implication, and **R3** states that a revelation of a revelation amounts to a revelation. (Imagine someone telling A that B was told that C's credit card pin number is C's birth date.) **BR1** and **BR2** demonstrate the intimate relation between belief and revelation. **BR1** is a weakened variant of a K axiom for R; the requirement that $\neg\phi$ is not believed is necessary to avoid cases where $\phi \rightarrow \psi$ is only

[6] Constraints on the semantic structure that ensure the validity of these axioms were identified but are not discussed here for limitations of space. Most of these are standard, however, except possibly for those pertaining to the revelation axioms.

$KD45$ axioms for B.
KD axioms for I.

IB **bridge axioms.**
IB1. $\neg I(x,\phi) \leftrightarrow B(x,\neg I(x,\phi))$
IB2. $I(x,\phi) \leftrightarrow B(x,I(x,\phi))$
IB3. $I(x,\phi) \rightarrow \neg B(x,\neg\phi)$

R **axioms.**
R1. $R(x,\phi) \wedge \neg R(x,\neg\phi)$, if $\vdash \phi$
R2. $R(x,\phi) \rightarrow R(x,\psi)$, if $\vdash \phi \rightarrow \psi$
R3. $R(x,R(y,\phi)) \rightarrow R(x,\phi)$

BR **bridge axioms.**
BR1. $[B(x,\phi \rightarrow \psi) \wedge \neg B(x,\neg\phi)] \rightarrow [R(x,\phi) \rightarrow R(x,\psi)]$
BR2. $R(x,\phi) \rightarrow B(x,R(x,\phi))$

Group axioms.
G1. $Mem(x,[y]) \leftrightarrow x = y$
G2. $Mem(x,G1 \sqcup G2) \leftrightarrow Mem(x,G1) \vee Mem(x,G2)$
G3. $Mem(x,G1 \sqcap G2) \leftrightarrow Mem(x,G1) \wedge Mem(x,G2)]$

VEL Axioms [28].
TP1. $H(\phi,t)$, if $\vdash \phi$
TP2. $(t \leq t' \wedge t' \leq t'') \rightarrow t \leq t'$
TP3. $t \leq t' \vee t' \leq t$
TP4. $(t \leq t' \wedge t' \leq t) \leftrightarrow t = t'$
TP5. $(H(\phi,t) \wedge H(\phi \rightarrow \psi,t)) \rightarrow H(\psi,t)$
TP6. $\neg H(\phi \wedge \neg\phi,t)$
TP7. $H(\phi,t) \vee H(\neg\phi,t)$
TP8. $H(\phi,t) \leftrightarrow H(H(\phi,t),t')$
TP9. $t \leq t' \leftrightarrow H(t \leq t',t'')$
TP10. $\forall t[H(\phi,t')] \rightarrow H(\forall t[\phi],t')$
TP11. $AT(t) \wedge AT(t') \rightarrow t = t'$
TP12. $H(AT(t),t)$
TP13. $\phi \rightarrow \exists t[H(\phi,t)]$
BA1. $\circledast\phi$, if $\vdash \phi$
BA2. $(\circledast\phi \wedge \circledast(\phi \rightarrow \psi)) \rightarrow \circledast\psi$
BA3. $\circledast\phi \rightarrow \phi$
BA4. $AT(t) \rightarrow \circledast AT(t)$
BA5. $t \leq t' \rightarrow \circledast(t \leq t')$
BA6. $H(\circledast\phi,t) \wedge t \leq t' \rightarrow H(\circledast H(\phi,t),t')$

Figure 1. System Σ of $\mathscr{L}_S$ axioms

- **T1.** $R(x,\phi \wedge \psi) \rightarrow R(x,\phi) \wedge R(x,\psi)$
- **T2.** $B(x,\phi) \rightarrow R(x,\phi)$
- **T3.** $R(x,\phi) \leftrightarrow R(x,R(x,\phi))$
- **T4.** $B(x,\phi) \rightarrow B(x,R(x,\phi))$
- **T5.** $B(x,R(x,\phi)) \rightarrow R(x,\phi)$
- **T6.** $B(x,\neg R(x,\phi)) \rightarrow \neg R(x,\phi)$
- **T7.** $R(x,B(x,\phi)) \rightarrow B(x,R(x,\phi))$

Figure 2. Some theorems of Σ

trivially believed. **BR2** means that agents have complete beliefs about their revelations. The revelation theorems in Figure 2 can be easily proved to follow from Σ.[7]

Henceforth, we make use of the following abbreviation: If O is B, I, R, or Mem; α and β are terms of the appropriate sorts; and t is of sort σ_T then we write $O(\alpha,\beta,t)$ as a shorthand for $H(O(\alpha,\beta),t)$. The following definition is a precise characterization of the simplest, bare-bones notion of secrecy based on the intuitions presented in Section 2:

$$Secret_0(\phi,K,N,\psi,t) =_{\text{def}} \neg NFE(K,t) \wedge \forall x[Mem(x,K,t) \rightarrow B(x,\phi \wedge \psi \wedge \neg NFE(N,t),t) \wedge \neg\mathscr{B}(\phi,x,N,t) \wedge \mathscr{I}(\phi,x,N,\psi,t)]$$

where

$$NFE(G) =_{\text{def}} \circledast\neg\mathbf{F}(\exists x[Mem(x,G)])$$
$$\mathbf{F}\phi =_{\text{def}} \exists t1,t2[AT(t1) \wedge t1 \leq t2 \wedge H(\phi,t2)]$$
$$\mathscr{B}(\phi,\alpha,N,t) =_{\text{def}} B(\alpha,\ \exists y[Mem(y,N,t) \wedge R(y,\phi,t)],\ t)$$

[7] All proofs are available here.

$$\mathscr{I}\ (\phi,\alpha,N,\psi,t) =_{\text{def}} \forall y,t'[I(\alpha, t \leq t' \wedge Mem(y,N,t') \wedge \forall t''[t < t'' \leq t' \rightarrow H(\psi,t'')] \rightarrow \neg R(y,\phi,t'),\ t)]$$

Thus, at time t, group K keeps the secretum ϕ a secret from group N, under the condition ψ if, at t, the group of secret keepers is not necessarily forever empty and each secret keeper

1. believes ϕ, ψ and that the group of nescients is not necessarily forever empty (**I5**, **I8**);
2. does not believe that there is a nescient to whom ϕ is revealed at t (**I9**); and
3. has the intention that at all future times t', such that ψ persists from t through t', ϕ is not revealed to any nescient (**I10**).

5. Seven Theorems on Secrets

In this section, we prove some results about secrets. Some of these are quite intuitive; others may seem counter-intuitive at first glance, but they are instructive in that they sharpen our intuitions about secrets. Henceforth, we write S to refer to the statement $Secret_0(\phi,K,N,\psi,t)$.

First, it should be uncontroversial that a revelation of the secrecy of ϕ is a revelation of ϕ (at a time when there is at least one keeper). Consequently, a keeper does not believe that there is a nescient to whom the secrecy of ϕ is revealed.

Theorem 1 *The following statements follow from* Σ.

1. $R(x,S \wedge \exists y Mem(y,K,t),t) \rightarrow R(x,\phi,t)$
2. $S \wedge Mem(x,K,t) \rightarrow \neg B(x,\exists y[Mem(y,N,t) \wedge R(y,S \wedge \exists z Mem(z,K,t),t)],t)$

Beliefs, intentions and revelations of a group g are inherited by every *subgroup* thereof. Where $g \sqsubseteq g' =_{\text{def}} \forall t,x[Mem(x,g,t) \rightarrow Mem(x,g',t)]$, the following follows.

Lemma 1 *The following statements are entailed by* Σ.

1. $(\forall x,t[Mem(x,g',t) \rightarrow B(x,\phi,t)] \wedge g \sqsubseteq g') \rightarrow \forall y,t[Mem(y,g,t) \rightarrow B(y,\phi,t)]$
2. $(\forall x,t[Mem(x,g',t) \rightarrow I(x,\phi,t)] \wedge g \sqsubseteq g') \rightarrow \forall y,t[Mem(y,g,t) \rightarrow I(y,\phi,t)]$
3. $(\forall x,t[Mem(x,g',t) \rightarrow R(x,\phi,t)] \wedge g \sqsubseteq g') \rightarrow \forall y,t[Mem(y,g,t) \rightarrow R(y,\phi,t)]$

Hence, secrets are also inherited by subgroups (assuming that the keepers' sub-group is not necessarily forever empty and every member of it believes that the nescients' sub-group is not necessarily forever empty). It follows that, given two secrets with the same secretum and secrecy condition, the intersection of the keepers is keeping the secret from the union of the nescients and the union of the keepers is keeping the secret from the intersection of the nescients.

Theorem 2 *The following are entailed by* Σ.

1. $S \wedge (K' \sqsubseteq K) \wedge (N' \sqsubseteq N) \wedge \neg NFE(K',t) \wedge \forall x[Mem(x,K',t) \rightarrow B(x,\neg NFE(N',t),t)] \rightarrow Secret_0(\phi,K',N',\psi,t)$

2. $Secret_0(\phi, K1, N1, \psi, t) \wedge Secret_0(\phi, K2, N2, \psi, t) \rightarrow$
$[\neg NFE(K1 \sqcap K2, t) \rightarrow Secret_0(\phi, K1 \sqcap K2, N1 \sqcup N2, \psi, t)] \wedge$
$[\forall x[Mem(x, K1 \sqcup K2, t) \rightarrow B(x, \neg NFE(N1 \sqcap N2, t), t)]$
$\rightarrow Secret_0(\phi, K1 \sqcup K2, N1 \sqcap N2, \psi, t)]$

Given that keepers are consistent believers, the secrecy condition ψ must, at any time t, be consistent with each keeper's beliefs and intentions, lest the group of keepers happens to be empty at t. Given that keepers believe the secretum and certain properties on the nescients and what is revealed to them the following theorem highlights some particularly important aspects of this constraint on the secrecy condition.

Theorem 3 *The following follows from* Σ.
$S \rightarrow \neg\exists x[Mem(x, K, t) \wedge B(x, \psi \rightarrow [\neg\phi \vee NFE(N, t) \vee \mathscr{B}(\phi, x, N, t) \vee \neg\mathscr{I}(\phi, x, N, \psi, t)], t)]$

The next theorem captures the intuition that secreta should not be bound to be revealed to the nescients while the secrecy condition holds. (This includes the trivial case where the secretum is a tautology.)

Theorem 4 *The following follows from* Σ.

$S \rightarrow \neg\exists x[Mem(x, K, t) \wedge B(x, \exists y, t'[t < t' \wedge Mem(y, N, t') \wedge$
$\forall t''[t < t'' \leq t' \rightarrow H(\psi, t'')] \wedge R(y, \phi, t')], t)]$

The clauses of the following theorem indicate that, given $Secret_0(\phi, K, N, \psi, t)$, under certain conditions some propositions, other than ϕ, are also secreta or are believed to be secreta by members of K.

Theorem 5 *The following statements follow from* Σ.

1. $S \wedge Secret_0(\xi, K, N, \psi, t) \rightarrow Secret_0(\phi \wedge \xi, K, N, \psi, t)$
2. $S \rightarrow Secret_0(\exists x R(x, \phi, t), K, N, \psi, t)$
3. $B(x, S \wedge Mem(x, K, t), t) \rightarrow Secret_0(\phi, [x], N, \psi, t)$
4. $S \wedge Mem(x, K, t) \rightarrow B(x, Secret_0(\phi, [x], N, \psi, t), t)$
5. $S \wedge \forall x[Mem(x, K, t) \rightarrow B(x, S \wedge \exists y[Mem(y, K, t)], t)] \rightarrow$
$Secret_0(S \wedge \exists y[Mem(y, K, t)], K, N, \psi, t)$
6. $S \wedge Mem(x, K, t) \rightarrow B(x, Secret_0(Secret_0(\phi, [x], N, \psi, t), [x], N, \psi, t), t)$

The first two clauses should be obvious enough: the conjunction of two secreta is a secretum and so is the revelation of a secretum to some agent. According to the third clause, an agent who believes that there is a secret of some group, and that they are a member of that group, is actually holding the secret. This is so even though the agent may be mistaken about the group's holding the secret or about their membership in the group. Clause 4 indicates that each secret keeper believes that the secret is kept by the group to which only they belong. This is the closest we can get to an *introspection* result for secrets; in particular, this keeper may be keeping the secret but is not aware of the existence of the group or of its keeping the secret. However, as per the fifth clause, if every secret keeper is aware of the existence of the group and of its keeping the secret, then the secrecy of the secretum is itself a secretum of the same group of keepers from the same group of nescients under the same secrecy condition. Nevertheless, by Clause 6, even in

this case where the keepers are aware of the group secret, they might still not believe in the secrecy of the secret for the group, simply because they may fail to believe that other keepers are aware of the secret. Hence, we can only prove a result akin to Clause 4.[8]

The following theorem presents *separation* results about K and N. First, no secret keeper believes that they are a nescient (Clause 1). Hence, no keeper is a nescient if the identity of nescients is known by each keeper (Clause 2). On the other hand, it may happen that an agent A who was once a secret keeper becomes a nescient. (Imagine players of team A keeping a secret from team B and at some later time an A player joins team B) This, however, does not mean that the secret is no longer kept; there are at least three reasons for this. First, the current secret keepers may not be aware of this conversion of their old co-keeper; second, they may not be aware that A *was* a co-keeper; and, third, it may be the case that the secretum, though once believed by A, is no longer revealed to them. While this last possibility is indeed moot, we do not want to commit to the permanence of revelation. However, assuming that a secret keeper is aware of the relevant facts and believes in the persistence of revelation, a contradiction is inevitable (Clause 3, where S' is just like S with t replaced by t'.)

Theorem 6 *The following follow from* Σ.

1. $S \wedge Mem(x,K,t) \rightarrow \neg B(x,Mem(x,N,t),t)$
2. $S \wedge Mem(x,K,t) \wedge \forall y[Mem(y,N,t) \rightarrow B(x,Mem(y,N,t),t)] \rightarrow \neg Mem(x,N,t)$
3. $[S \wedge Mem(x,K,t') \wedge B(x,t \leq t' \wedge [R(C,\phi,t) \rightarrow R(C,\phi,t')] \wedge Mem(x,K,t')$
   $\wedge S \wedge Mem(C,K,t) \wedge Mem(C,N,t'),t') \rightarrow B(x,\neg S',t')]$

Finally, if a secret keeper, A, believes that nescients believe that ξ implies the secretum ϕ, then A does not believe that ξ is revealed to any nescient and they do not intend to reveal it as long as the secrecy condition holds. Note, however, that ξ need not be a secretum since it is possible that A does not believe it.

Theorem 7 *The follows follows from* Σ.
$S \wedge Mem(x,K,t) \wedge B(x,\forall y[Mem(y,N,t) \rightarrow \neg B(y,\neg\xi,t) \wedge B(y,\xi \rightarrow \phi,t)],t) \rightarrow$
$\neg\mathscr{B}(\xi,x,N,t) \wedge$
$\neg\mathscr{I}(x,\forall y,t'[t \leq t' \wedge Mem(y,N,t') \wedge \forall t''[t \leq t'' \leq t' \rightarrow H(\psi,t'')] \rightarrow R(y,\xi,t')],t)$

6. A Typology of secrets

$Secret_0$ is only a bare-bones and, hence, weak notion of secrecy. We intuitively think of most secrets as involving stronger conditions. Five such stronger notions of secrecy are shown in Figure 3; all imply the bare-bones notion. $Secret_1$ is a secret in which the secretum is indeed not revealed to the nescients; $Secret_2$ is a secret of which keepers are aware; and $Secret_3$ holds when the keepers believe that the secretum ϕ is indeed not revealed to the nescients. In a $Secret_4$ situation, keepers are aware of the identity of all keepers, while, in a $Secret_5$ situation, they are aware of their membership in the group and believe that all keepers are aware of the secret.

[8] For most common cases of secrecy, stronger results can be proven since, in such cases, keepers are typically aware of the existence of K and of their membership thereof. In particular, we can prove that if the condition of membership in K is mere revelation of the secretum (which is typical of many secrets) each keeper believes that they are a member of K: $S \wedge Mem(x,K,t) \wedge B(x,\forall y[Mem(y,K,t) \leftrightarrow R(y,\phi,t)],t) \rightarrow B(x,Mem(x,K,t),t)$.

1. $Secret_1(\phi,K,N,\psi,t) =_{def} S \wedge \forall y[Mem(y,N,t) \rightarrow \neg R(y,\phi,t)]$
2. $Secret_2(\phi,K,N,\psi,t) =_{def} S \wedge \forall x[Mem(x,K,t) \rightarrow B(x,S,t)]$
3. $Secret_3(\phi,K,N,\psi,t) =_{def}$
   $S \wedge \forall x[Mem(x,K,t) \rightarrow B(x,\forall y[Mem(y,N,t) \rightarrow \neg R(y,\phi,t)],t)]$
4. $Secret_4(\phi,K,N,\psi,t) =_{def}$
   $S \wedge \forall x,y[Mem(x,K,t) \rightarrow [Mem(y,K,t) \leftrightarrow B(x,Mem(y,K,t),t)]]$
5. $Secret_5(\phi,K,N,\psi,t) =_{def}$
   $S \wedge \forall x[Mem(x,K,t) \rightarrow B(x,Mem(x,K,t) \wedge \forall y[Mem(y,K,t) \rightarrow B(y,S,t)],t)]$

Figure 3. Some common stronger notions of secrecy

Henceforth, we write S_n where is n is 1,2,3,4 or 5 referring to the corresponding secret type. Perhaps most common secrets are instances of all five types, satisfying $\bigwedge_{i=1}^{5} S_i$. These types are not totally independent though, as demonstrated by the following theorem. The first clause states that S_2 and S_3 hold if and only if there is a secret and every secret keeper believes the secret and that the secretum is not revealed to any nescient (S_1). By Clause 2, S_2 follows immediately from S_5. If all the keepers are unmistakably aware of one another (S_4) then S_5 holds if and only if all keepers believe both the secret and that all co-keepers believe the secret (S_2). Of particular interest is the fourth clause which indicates that S is equivalent to S_3 in case the secrecy condition implies (or *is*) that the secretum is not revealed to a nescient, which is a quite common condition of secrecy.

Theorem 8

1. $\Sigma \vdash S_2 \wedge S_3 \leftrightarrow S \wedge \forall x[Mem(x,K,t) \rightarrow B(x,S_1,t)]$
2. $\Sigma \vdash S_5 \rightarrow S_2$
3. $\Sigma \vdash S_4 \rightarrow [S_5 \leftrightarrow \forall x[Mem(x,K,t) \rightarrow B(x,S_2,t)]]$
4. If $\psi \vdash \forall y[Mem(y,N,t) \rightarrow \neg R(y,\phi,t)]$, *then* $\Sigma \vdash S \rightarrow S_3$

The clauses of Theorem 9 state that, depending on the secret type, secret keepers are bound to have certain properties (mostly beliefs). In an S_1 situation, we get complete separation of the groups of keepers and nescients (Clause 1); this separation is only a belief of each keeper in an S_2 situation (Clause 2). Similar results are indicated by Clauses 3 and 4 but with respect to the nescients' not believing the secret. The fifth clause states that, given S_2, every keeper holds the *de dicto* belief that members of K are individually keeping the secret. The sixth clause states that the same belief is held *de re* if both S_2 and S_4 hold.

Theorem 9 *The following follow from* Σ.

1. $S_1 \rightarrow [Mem(x,K,t) \rightarrow [\neg Mem(x,N,t)]]$
2. $S_2 \rightarrow [Mem(x,K,t) \rightarrow B(x,\neg\exists y[Mem(y,K,t) \wedge Mem(y,N,t)],t)]$
3. $S_1 \rightarrow [\forall y[Mem(y,N,t) \rightarrow \neg B(y,S_0 \wedge \exists z Mem(z,K,t),t)]]$
4. $S_3 \rightarrow [Mem(x,K,t) \rightarrow B(x,\forall y[Mem(y,N,t) \rightarrow \neg B(y,S_0 \wedge \exists z[Mem(z,K,t)],t)],t)]$
5. $S_2 \rightarrow [Mem(x,K,t) \rightarrow B(x,[Mem(y,K,t) \rightarrow Secret_0(\phi,[y],N,\psi,t)],t)]$
6. $S_4 \wedge S_2 \rightarrow [Mem(x,K,t) \rightarrow [Mem(y,K,t) \rightarrow B(x,Secret_0(\phi,[y],N,\psi,t),t)]]$

7. Related Work

Philosophical and psychological investigations of secrets are best represented by the work of Bok in philosophy [1] and Kelly's book [2] and the work of Slepian et al [3,22, for example] in psychology. These authors share our intuition that secrecy is mostly about the intention to conceal. Their interests in secrets are different from ours though; they are primarily interested in ethical issues related to secrets [1] and in the motivations for and the psychological effects of keeping secrets [2,3,22]

Logical accounts of secrecy abound in the literature on system security [5,6,7,8,9, 10,11,12,13,14, for instance]. Much of this literature is rooted in, or best represented by, the work of Halpern and O'Neill [9]. The authors consider multi-agent systems with a branching time structure where, at any time, each agent is in some *local state* comprising all the information accessible to the them. Agents are never mistaken about their local states; they never hold false beliefs. Using this machinery, Halpern and O'Neill define several notions of secrecy. The most fundamental of these, *total secrecy*, is defined as follows. The actual local state of agent j is totally secret from agent i if i cannot "rule out" any possible local state of j.

The usefulness of this account, and of most other accounts in the literature [11,13, for example], is based on a couple of assumptions:

1. **Local states are typically a collection of assignments of values to variables.** If said variables correspond to propositions, then, assuming a classical bivalent logic, there can only be two values: true and false. Hence, not being able to rule out a local state amounts to not being able to decide whether a proposition is true or false.
2. **Agents cannot hold false beliefs.** As pointed out above, this is built into the theory. Hence, given the first point, if we think of secrets as propositions, a proposition can only be a secret from agent i if i is in suspense about the proposition.
3. **Systems can be constructed.** The assumption here is that it is always possible to fully characterize the system as a branching tree of states. This is probably always possible if systems are programs or simple database transactions [11, for example].

Given our objective to characterize secrecy in an unconstrained, commonsense setting, we cannot uphold any of the above assumptions. First, since we consider objects of secrecy to be only propositions, assumption 1 reduces to the case of variables with binary domains. Second, we cannot in general make the unrealistic assumption that agents have no false beliefs; assumption 2 does not allow us to account for situations where j keeps P a secret from i who believes $\neg P$. Third, in a general theory of secrets allowing all forms of complex social interactions, the assumption of a system which is constructible as a branching tree of states is at least questionable.

The revelation modality we introduced does not seem to have a thoroughly investigated precedent. It is perhaps possible to use a variant of the notion of *announcement* from dynamic epistemic logic [30,31] to model *acts* of revealing. This would, however, require, possibly extensive, revision of the principles underlying the logic of announcement. In particular, a *truthful announcement* of P results in the addressee's believing P (at least if P is atomic) and a *possibly lying announcement* thereof causes the addressee to believe that the announcer believes P [30,31]. Even if we adopt the latter, more cautious attitude towards announcement as a model of revelation, we are restricted to revelations made only by cognitive agents which may have beliefs. Our passive notion of revelation does not

require this and is consistent with a revelation resulting from a simple act of perception not involving an announcing agent.

8. Conclusion

We presented foundations for a logical, commonsense theory of secrets. A secret is construed as a situation in which a group of secret keepers believe a proposition, which they do not believe to have been revealed to members of another group of nescients. Crucially, the keepers intend that this concealment from the nesciencts persists so long as some condition of secrecy holds. To that end, a non-normal modal operator for revelation was identified together with axioms relating it to belief. Further, towards an ontology of secrets, various types of secrets were identified, all of which including the bare-bones definition in addition to some extra common conditions. Several properties which sharpen our intuitions about secrets were proven and more are to be investigated in future work.

9. Acknowledgment

We are thankful to our colleagues Nourhan Ehab, Nada Sharaf and Ammar Yasser for their comments on earlier versions of the paper. We would also like to express our deepest gratitude to the reviewers for their helpful suggestions.

References

[1] Bok S. Secrets: On the Ethics of Concealment and Revelation. Vintage; 1989.
[2] Kelly AE. The psychology of secrets. Springer Science & Business Media; 2002.
[3] Slepian ML, Chun JS, Mason MF. The experience of secrecy. Journal of Personality and Social Psychology. 2017;113(1):1.
[4] Vincent J, Taipale S, Sapio B, Lugano G, Fortunati L, editors. Social Robots from a Human Perspective. Switzerland: Springer; 2015.
[5] Bieber P. A logic of communication in hostile environments. In: Proceedings of the Computer Security Foundations Workshop III; 1990. p. 14–22.
[6] Glasgow J, MacEwen G, Panangaden P. A logic for reasoning about security. In: Proceedings of the Computer Security Foundations Workshop III; 1990. p. 2–13.
[7] Cuppens F. A logical formalization of secrecy. In: Proceedings of the Computer Security Foundations Workshop VI; 1993. p. 53–62.
[8] Roy A, Datta A, Derek A, Mitchell JC, Seifert JP. Secrecy analysis in protocol composition logic. In: Annual Asian Computing Science Conference. Springer; 2006. p. 197–213.
[9] Halpern JY, O'Neill KR. Secrecy in multiagent systems. ACM Transactions on Information and System Security (TISSEC). 2008;12(1):5:1–5:47.
[10] More SM, Naumov P. An Independence Relation for Sets of Secrets. Studia Logica. 2010;94(1):73–85.
[11] Biskup J, Tadros C. Preserving confidentiality while reacting on iterated queries and belief revisions. Annals of Mathematics and Artificial Intelligence. 2015;73:75–123.
[12] Tao J, Slutzki G, Honavar V. A Conceptual Framework for Secrecy-preserving Reasoning in Knowledge Bases. ACM Transactions on Computational Logic. 2014;16(1):3:1–3:32.
[13] Tsukada Y, Sakurada H, Mano K, Manabe Y. On compositional reasoning about anonymity and privacy in epistemic logic. Annals of Mathematics and Artificial Intelligence. 2016;78:101–129.
[14] Ramezanifarkhani T, Owe O, Tokas S. A secrecy-preserving language for distributed and object-oriented systems. Journal of Logical and Algebraic Methods in Programming. 2018;99:1–25.
[15] Marshall D, Weatherson B. Intrinsic vs. Extrinsic Properties. In: Zalta EN, editor. The Stanford Encyclopedia of Philosophy. spring 2018 ed. Metaphysics Research Lab, Stanford University; 2018. .
[16] Link G. Algebraic Semantics in Language and Philosophy. Stanford, CA: CSLI Publications; 1998.
[17] Landman F. Groups I. Linguistics and Philosophy. 1989;12(5):559–605.
[18] Galton A, Wood Z. Extensional and intensional collectives and the *de re/de dicto* distinction. Applied Ontology. 2016;11(3):205–226.
[19] Ismail HO. Stability in a Commonsense Ontology of States. In: Proceedings of the Eleventh International Symposium on Commonsense Reasoning (Commonsense-2013). Agya Napa, Cyprus; 2013. .

[20] The United States Department of Justice. Declassification; 2018. https://www.justice.gov/archives/open/declassification.
[21] Lane JD, Wegner DM. The cognitive consequences of secrecy. Journal of personality and social psychology. 1995;69(2):237.
[22] Slepian ML, Greenaway KH. The benefits and burdens of keeping others' secrets. Journal of Experimental Social Psychology. 2018;78:220–232.
[23] Mahon JE. The Definition of Lying and Deception. In: Zalta EN, editor. The Stanford Encyclopedia of Philosophy. winter 2016 ed. Metaphysics Research Lab, Stanford University; 2016. .
[24] Cohen PR, Levesque HJ. Intention is choice with commitment. Artificial intelligence. 1990;42(2-3):213–261.
[25] Beaver DI. Presupposition and Assertion in Dynamic Semantics. Stanford, CA: CSLI Publications; 2011.
[26] Fagin R, Halpern J. Belief, awareness, and limited reasoning. Artificial Intelligence. 1987;34(1):39–76.
[27] van Benthem J, Fernandez-Duque D, Pacuit E. Evidence and plausibility in neighborhood structures. Annals of Pure and Applied Logic. 2014;165(1):106–133.
[28] Bennett B, Galton AP. A unifying semantics for time and events. Artificial Intelligence. 2004;153(1-2):13–48.
[29] Pacuit E. Neighborhood semantics for modal logic. Springer; 2017.
[30] van Ditmarsch H, van der Hoek W, Kooi B. Dynamic Epistemic Logic. Dordrecht: Springer; 2008.
[31] Van Ditmarsch H. Dynamics of lying. Synthese. 2014;191(5):745–777.