Optimal Eventual Byzantine Agreement Protocols with Omission Failures

KAYA ALPTURER, Cornell University, USA
JOSEPH Y. HALPERN, Cornell University, USA
RON VAN DER MEYDEN, UNSW Sydney, Australia

Work on *optimal* protocols for *Eventual Byzantine Agreement* (EBA)—protocols that, in a precise sense, decide as soon as possible in every run and guarantee that all nonfaulty agents decide on the same value—has focused on *full-information protocols* (FIPs), where agents repeatedly send messages that completely describe their past observations to every other agent. While it can be shown that, without loss of generality, we can take an optimal protocol to be an FIP, full information exchange is impractical to implement for many applications due to the required message size. We separate protocols into two parts, the *information-exchange protocol* and the *action protocol*, so as to be able to examine the effects of more limited information exchange. We then define a notion of optimality with respect to an information-exchange protocol. Roughly speaking, an action protocol P is optimal with respect to an information-exchange protocol $\mathcal{E}$ if, with P, agents decide as soon as possible among action protocols that exchange information according to $\mathcal{E}$. We present a knowledge-based EBA program for omission failures all of whose implementations are guaranteed to be correct and are optimal if the information exchange satisfies a certain safety condition. We then construct concrete programs that implement this knowledge-based program in two settings of interest that are shown to satisfy the safety condition. Finally, we show that a small modification of our program results in an FIP that is both optimal and efficiently implementable, settling an open problem posed by Halpern, Moses, and Waarts (SIAM J. Comput., 2001).

CCS Concepts: • **Theory of computation** → **Distributed algorithms**; • **Computing methodologies** → *Reasoning about belief and knowledge*; • **Computer systems organization** → **Dependable and fault-tolerant systems and networks**.

Additional Key Words and Phrases: Distributed algorithms, Epistemic logic, Reasoning about knowledge, Byzantine Agreement, Consensus, Fault tolerance

1 INTRODUCTION

Logics of knowledge (epistemic logics) have been shown to provide useful abstractions for reasoning about distributed systems [5, 7], enabling a focus on the information that needs to be attained in order to perform certain actions, independent of how that information is encoded in the states of the system. The approach has, in particular, been used fruitfully to analyze agreement protocols, where agents are required to make consistent decisions on some value, based on their initial preferences [1, 4, 8, 12].

One particular focus of this work has been to develop protocols that make optimal use of information. In the context of protocols for agreement, this has meant protocols that decide as soon as possible. We say that a protocol P_1 *dominates* a protocol P_2 if, roughly speaking, for each possible pattern of failures and inputs, P_1 decides at least as soon as P_2. An *optimum protocol* is one that dominates every other protocol; an *optimal protocol* is one that is not dominated by any other protocol. Optimum knowledge-based programs (i.e., programs with explicit tests for knowledge) have been provided for simultaneous Byzantine agreement (SBA) with crash failures

Authors' addresses: Kaya Alpturer, ki78@cornell.edu, Cornell University, Ithaca, NY, USA; Joseph Y. Halpern, halpern@cs.cornell.edu, Cornell University, Ithaca, NY, USA; Ron van der Meyden, R.VanderMeyden@unsw.edu.au, UNSW Sydney, Sydney, NSW, Australia.

[1, 2, 4] and omission failures[1] [12], and for *consistent* SBA with omission failures [13] (where all agents that decide on an action must perform the same action, not just the nonfaulty agents). Moreover, polynomial-time implementations of these programs were provided. For the problem of eventual Byzantine agreement (EBA), it is well-known that there are no optimum protocols [12], although there are optimal protocols. Optimal knowledge-based programs have been provided for EBA in the case of crash failures [1] and sending-omission failures [8]. While a polynomial-time implementation of the knowledge-based program was provided for the case of crash failures [1], none was provided in the case of omission falures. Indeed, the problem of finding a polynomial-time optimal algorithm for EBA in the presence of omission failures has been open for over 20 years. Among other things, we solve this problem here.

The work on optimal (and optimum) protocols has focused on *full-information protocols* (FIPs), ones where each agent repeatedly sends all other agents its complete state, containing its initial state and all messages that it has received up to the present time. As far as optimal protocols go, there is no loss of generality in considering only FIPs. As is well known [3], any protocol can be simulated by an FIP, so for any protocol, there is an FIP that decides at least as soon. However, while FIPs do provide optimality, they are expensive to implement due to their space requirements and the length of messages sent, their analysis may be complex, and in some failure environments, they may require that intractable properties be computed at each step of the protocol to attain optimality. They are therefore not necessarily practical. The present paper is part of a program of research in which we seek to overcome these difficulties with the full-information paradigm by considering protocols in which less than full information is exchanged between the agents.

Our goal in this paper is to examine the effects of more limited information exchange. In order to do so, we separate protocols into two parts, the *information-exchange protocol*, which specifies what information agents maintain in their local states and what message they exchange at each step, and the *action protocol*, which, in the case of agreement protocols, specifies what decision agents make. We then define a notion of optimality with respect to an information-exchange protocol. Roughly speaking, an action protocol P is optimal for a particular specification with respect to an information-exchange protocol $\mathcal{E}$ if agents decide as soon as possible with P as they do with any other protocol that satisfies the specification and exchanges information according to $\mathcal{E}$. A full-information protocol is a special case of an information-exchange protocol, but we allow far more limited protocols, where agents keep track of less information and send less information in their messages. We focus in this paper on optimal protocols for binary EBA, where agents have only one of two possible initial preferences, and we assume sending-omission failures. For EBA, there is a group of agents, each with an initial preference of 0 or 1. The nonfaulty agents are required to reach agreement on a value, but different agents may decide at different times. The solution is required to be non-trivial in the sense that if all initial preferences are $x \in \{0, 1\}$ then a nonfaulty agent must decide x.

Since with EBA, agents do not have to decide simultaneously, the literature has typically considered optimal protocols that are biased towards 0, in that an agent decides 0 as soon as it learns that some agent started with an initial preference of 0, and protocols that are baised towards 1. In the case of crash failures, it is known that there are optimal EBA protocols that are biased towards 0 (resp., 1) in this sense [1]. But it is easy to see that there cannot be an optimal EBA protocol that is biased towards 0 (or 1) in the case of omissions failures. Consider a setting where there are at least three agents, and a run r where exactly one of these agents is faulty, say agent 1, all the remaining

[1] Recall that with *crash failures*, a faulty process behaves according to the protocol, except that it might crash at some round (possibly after sending some messages), after which it sends no messages; with sending-omission failures, a faulty process may omit to send an arbitrary set of messages in any given round, but otherwise behaves according to the protocol.

agents have an initial preference of 1, and the faulty agent does not send any messages. In run r, the nonfaulty agents must eventually decide 1, because the faulty agent's initial preference may be 1, and in that case, EBA requires a decision of 1. Suppose that the first nonfaulty agent to decide in run r does so at round n, and without loss of generality, that 2 is a nonfaulty agent that decides 1 at round n. Now consider a run r' where agent 1 has an initial preference of 0, agent 1 is faulty, all the remaining agents are nonfaulty and have an initial preference of 1, agent 1 does not send any messages up to round $n-1$, and in round $n-1$ sends exactly one message, which is sent to agent 3, and says (truthfully) that agent 1's initial preference was 0. Since agent 2 cannot distinguish r from r', agent 2 must decide 1 in round n of r'. Since agent 3 does not decide in the first $n-1$ rounds of r, it also does not decide in the first $n-1$ rounds of r'. Since it learns that some agent has an initial preference of 0 in round $n-1$, agent 3 decides 0 in round n of r'. Thus, two nonfaulty agents decide on different values in r', so EBA is not achieved. (Note a run like r' is inconsistent with crash failures; it really requires omission failures.)

To deal with this issue, in a 0-biased protocol, rather than requiring that an agent decide 0 as soon as it hears about a 0, we require only that an agent decides 0 only if it hears about 0 via a chain of agents (where the first agent in the chain has an initial preference of 0, and in round k, the kth agent in the chain decides 0 and tells the $(k+1)$st agent about this). Note that in the case of crash failures, an agent can hear about a 0 only via such a chain. We then provide a knowledge-based action protocol $\mathbf{P}^0$ based on this (well-known) idea that we show is correct (in the sense that all of its implementations satisfy the EBA specification, no matter what information-exchange protocol is used) and is optimal in contexts that satisfy a certain safety condition. We then consider two information-exchange protocols where agents exchange relatively short (and relatively few) messages, and show that they satisfy the safety condition. Finally, we provide concrete polynomial-time action protocols that implement $\mathbf{P}^0$ with respect to these two information-exchange protocols.

The knowledge-based program $\mathbf{P}^0$ is not optimal in the case of full-information contexts, but, as we show, a small modification $\mathbf{P}^1$ of $\mathbf{P}^0$ is optimal. Moreover, $\mathbf{P}^0$ and $\mathbf{P}^1$ are equivalent in the two limited-information contexts that we considered, so $\mathbf{P}^1$ is also optimal in these contexts. Roughly speaking, $\mathbf{P}^1$ allows agents to decide if they get common knowledge of who the nonfaulty agents are. Such common knowledge can't be obtained in the limited-exchange contexts that we consider, but it can be obtained if agents use a full-information protocol and keep track of everything they have learned. The knowledge-based program $\mathbf{P}^1$ does not involve the (rather complicated) *continual common knowledge operator* used by Halpern, Moses, and Waarts; rather, it uses more standard knowledge and common knowledge operators. This allows us to provide a polynomial-time implementation of it.

The rest of this paper is organized as follows. Section 2 reviews the epistemic framework of [5]. In Section 3, we introduce the separation of information-exchange and action protocols along with a representation of the failure model. Section 4 defines knowledge-based programs. The specification of EBA and a formal definition of optimality with respect to an information-exchange protocol are given in Section 5. In Section 6, we define the knowledge-based program $\mathbf{P}^0$, show that it satisfies EBA, define the safety condition that suffices for $\mathbf{P}^0$ to be optimal, describe two natural limited information-exchange protocols that satisfy the safety condition, and provide concrete action protocols that implement $\mathbf{P}^0$ with respect to these two information-exchange protocols. In Section 7, we define $\mathbf{P}^1$, a modification of $\mathbf{P}^0$ that is optimal for the full-information-exchange protocol, and provide a polynomial-time implementation of it. We conclude with a discussion of the cost of limited information exchange in Section 8. Proofs for all the results in the paper can be found in the appendix.

2 SEMANTIC MODEL

We assume that a set Agt of agents communicate using a message-passing network, which may be subject to various types of failures. To model such systems semantically, we use the standard runs-and-systems model [5, 6], which we briefly review.

Interpreted systems [5] model multi-agent scenarios in which some number n of agents change their states over time. An interpreted system is a pair $\mathcal{I} = (\mathcal{R}, \pi)$, where $\mathcal{R}$ is a set of runs, describing how the system evolves over time, and $\pi : \mathcal{R} \times \mathbb{N} \rightarrow \mathcal{P}(Prop)$ is an interpretation function that indicates which atomic propositions are true at each *point* of the system, where a point is a pair (r, m) consisting of a run $r \in \mathcal{R}$ and time $m \in \mathbb{N}$. The set $\mathcal{R}$ is called a *system*. Formally, a run $r \in \mathcal{R}$ is a function $r : \mathbb{N} \rightarrow L_e \times \Pi_{i \in Agt} L_i$, where L_e is the set of possible *local states* of the environment in which the agents operate, and each L_i is the set of possible *local states* of agent i. The elements of $L_e \times \Pi_{i \in Agt} L_i$ are called *global states*. Given a run r, agent i, and time m, we write $r_i(m)$ for the $(i+1)$st component of $r(m)$, which is the local state of agent i in the global state $r(m)$, and $r_e(m)$ for the first component of $r(m)$, which is the local state of the environment.

To reason about the knowledge of agents in interpreted systems, we use a standard language for reasoning about knowledge and time. We start with a set Φ of primitive propositions, and close off under $\wedge$, $\neg$, the epistemic operators K_i for $i = 1, \ldots, n$ (one for each agent) and $C_{\mathcal{S}}$ (common knowledge among the agents in an *indexical* set S; see below) and the temporal operators $\Box$, $\boxdot$, $\bigcirc$, and $\ominus$. The formula $K_i\phi$ says that agent i knows that formula ϕ holds, $\Box\phi$ says that ϕ holds at all times in the future, $\boxdot\phi$ says that ϕ holds at all times, $\bigcirc\phi$ says that ϕ holds at the next time, and $\ominus\phi$ says that ϕ holds at the previous time.

The semantics of the logic is given by a relation $\mathcal{I}, (r, m) \models \phi$, where $\mathcal{I}$ is an interpreted system, (r, m) is a point of $\mathcal{I}$, and ϕ is a formula. For formulas not of the form $C_{\mathcal{S}}\phi$, the relation $\models$ is defined inductively as follows (we omit the obvious cases for the propositional operators):

- $\mathcal{I}, (r, m) \models p$ if $p \in \pi(r, m)$,
- $\mathcal{I}, (r, m) \models K_i\phi$ if $\mathcal{I}, (r', m') \models \phi$ for all points (r', m') of $\mathcal{I}$ such that $r_i(m) = r'_i(m')$,
- $\mathcal{I}, (r, m) \models \Box\phi$ if $\mathcal{I}, (r, m') \models \phi$ for all $m' \geq m$,
- $\mathcal{I}, (r, m) \models \boxdot\phi$ if $\mathcal{I}, (r, m') \models \phi$ for all $m' \in \mathbb{N}$,
- $\mathcal{I}, (r, m) \models \bigcirc\phi$ if $\mathcal{I}, (r, m+1) \models \phi$
- $\mathcal{I}, (r, m) \models \ominus\phi$ if $m > 0$ and $\mathcal{I}, (r, m-1) \models \phi$.

The intuition for the definition of the knowledge operator $K_i\phi$ is that $r'_i(m) = r_i(m)$ says that agent i considers it possible, when in the actual situation (r, m), that it is in situation (r', m'), since it has the same local state there. An agent then *knows* ϕ if ϕ is true in all the situations that the agent considers to be possible.

We can now define the modal operator $C_{\mathcal{S}}$. Intuitively, $C_{\mathcal{S}}\phi$ is true at a point (r, m) if ϕ is common knowledge among the agents in S; each of the agents in $\mathcal{S}$ knows that each of the agents in S knows ... that ϕ is true. The fact that $\mathcal{S}$ is an indexical set means that its membership can depend on the point; that is, semantically, $\mathcal{S}(r, m)$ is a set of agents for each point (r, m). We define $E_{\mathcal{S}}\phi$ (everyone in $\mathcal{S}$ knows ϕ) as an abbreviation for $\bigwedge_{i \in \mathcal{S}} K_i\phi$. That is,

$$\mathcal{I}, (r, m) \models E_{\mathcal{S}}\phi \text{ if, for all } j \in \mathcal{S}(r, m) \text{, we have } \mathcal{I}, (r, m) \models K_j\phi.$$

Taking $E^1_{\mathcal{S}}\phi$ to be an abbreviation of $E_{\mathcal{S}}\phi$, and $E^{m+1}_{\mathcal{S}}\phi$ to be an abbreviation of $E_{\mathcal{S}}(E^m_{\mathcal{S}}\phi)$, we define

$$\mathcal{I}, (r, m) \models C_{\mathcal{S}}\phi \text{ if, for all } m \geq 1, \mathcal{I}, (r, m) \models E^m_{\mathcal{S}}\phi.$$

As usual, we say that ϕ is *valid in* $\mathcal{I}$, and write $\mathcal{I} \models \phi$, if $\mathcal{I}, (r, m) \models \phi$ for all points (r, m) in $\mathcal{I}$.

3 COMMUNICATION AND FAILURE MODELS

We now specialize the general model from the previous section to represent an omissions-failure model. In our representation, we separate the *information-exchange protocol*, which characterizes the information maintained by agents in their local states, and which messages are sent and when, from the *action protocol*, which characterizes the rules for performing actions other than sending messages. (In our case, these actions will be decisions.) In the literature, the information-exchange protocol has often been the *full-information protocol*, in which at each step each agent sends all other agents a complete description of everything it has learned up to that time. However, we will be interested in protocols in which less information is exchanged, so it helps to separate out the information-exchange protocol as a parameter of the interpreted systems we construct. A further parameter is the failure model $\mathcal{F}$, defined below.

We assume that information is exchanged by sending messages. Our focus will be on *synchronous* message passing, in which agents operate in a sequence of synchronized rounds. In each round, each agent performs some actions, sends a set of messages to the other agents, receives some of the messages from the other agents that were sent in the same round, and updates its state depending on these events. The information-exchange protocol describes the possible initial states of the agent (which may include information such as the agent's preference for the outcome of the consensus decision to be made), how it chooses the messages to send at each time, and how it updates its state in response to receiving messages.

We assume that each agent has a set A_i of actions that it can perform. In our applications, $A_i = \{\mathtt{decide}_i(x) \mid x \in \{0,1\}\} \cup \{\mathtt{noop}\}$, but in general, A_i can be arbitrary. Formally, an information-exchange protocol $\mathcal{E}$ for agents $Agt = \{1, \ldots, n\}$ is given by a tuple $\langle \mathcal{E}_1, \ldots, \mathcal{E}_n \rangle$ consisting of a local information-exchange protocol $\mathcal{E}_i$ for each each agent i. Each local information-exchange protocol $\mathcal{E}_i$ is a tuple $\langle L_i, I_i, A_i, M_i, \mu_i, \delta_i \rangle$, where

- L_i is a set of local states.
- $I_i \subseteq L_i$ is a set of initial states.
- M_i is a set of messages that can be sent by agent i.
- $\mu_i : L_i \times A_i \to \Pi_{j \in Agt}(M_i \cup \{\perp\})$ is a function mapping a local state s and an action a to the messages to be sent in the current round, one to each agent j. Intuitively, $\mu_i(s,a) = \sigma$ means that when action a is performed in state s, the information-exchange protocol transmits message σ_j to each agent j. Here $\sigma_j = \perp$ represents that no message is sent by i to j. Let $\mu_{ij}(s,a)$ denote the message that i sends to j in this tuple.
- $\delta_i : L_i \times A_i \times \Pi_{j \in Agt}(M_j \cup \{\perp\}) \to L_i$ is a function that updates the local state, given an action and a tuple $(m_1, \ldots, m_n)$ of messages $m_j \in M_j \cup \{\perp\}$ (where $m_j = \perp$ if i receives no message from j).

The failure model describes what failures can occur. Typically a failure model comes with a parameter t that indicates the maximum number agents that may be faulty. A *failure pattern*, or *adversary*, defines the failures that actually occur in a particular run consistent with the failure model. Formally, a failure pattern α is a pair $(\mathcal{N}, F)$, where $\mathcal{N} \subseteq Agt$ and F is a mapping $F : \mathbb{N} \times Agt \times Agt \to \{0,1\}$. Here $\mathcal{N}$ is the set of nonfaulty agents, and $F(m,i,j)$ describes whether the message sent from agent i to agent j in round $m+1$ is delivered. (If it is not delivered, we assume that the message $\perp$ is delivered instead.) A failure model is a set of failure patterns. The sending-omissions model $SO(t)$ for agents Agt is the set of all failure patterns $(\mathcal{N}, F)$ such that $|Agt - \mathcal{N}| \leq t$, so that there are at most t faulty agents, and for all $m \in \mathbb{N}$ and $j \in Agt$, if $F(m,i,j) = 0$ then $i \in Agt - \mathcal{N}$. The crash-failures model is the special case where if $F(m,i,j) = 0$ then $F(m',i,j') = 0$ for all $m' > m$ and agents j'.

An *action protocol* P for an information-exchange protocol $\mathcal{E}$, is a tuple $(P_1, \ldots, P_n)$ containing, for each agent $i = 1 \ldots n$, a *local action protocol* $P_i : L_i \longrightarrow A_i$ mapping the local states L_i for agent i in $\mathcal{E}$ to actions in A_i.

To connect these definitions to the semantic model of Section 2, we describe how an information-exchange protocol $\mathcal{E}$, a failure model $\mathcal{F}$, and an action protocol P determine a system $\mathcal{R}_{\mathcal{E},\mathcal{F},P}$. In this system, the set L_e of possible local states of the environment consists of the possible failure patterns. The local states L_i of the agent are the states of the information-exchange protocol for agent i. An *initial state* $(s_e, s_1, \ldots, s_n)$ is a global state, where $s_i \in I_i$ is an initial state of each agent i's information-exchange protocol. For each initial state, a run r with that initial state is uniquely determined by the information-exchange protocol $\mathcal{E}$, the failure model $\mathcal{F}$, and the action protocol P. In this run, the protocol $\mathcal{E}$, the failure pattern α, and P determine at each step, in order, what actions are taken, what messages are sent, and what messages are received. Each agent updates its local state depending on the actions taken and the messages received in the round. Formally, the global state $r(k+1) = (s'_e, s'_1, \ldots, s'_n)$ at time $k+1$ is determined from the global state $r(k) = (s_e, s_1, \ldots, s_n)$ at time k as follows:

- $s'_e = s_e$ (so the failure pattern remains unchanged throughout the run).
- For each pair of agents i and j, let $m_{i,j}$ be the message that agent i sends to j, given that it performs action $P_i(s_i)$ in state s_i, that is, $m_{i,j} = \mu_i(s_i, P_i(s_i))(j)$.
- For each pair of agents i and j, let $m'_{i,j}$ be the result of applying the failure pattern to the messages sent. Specifically, suppose that $s_e = (\mathcal{N}, F)$. If $F(k, i, j) = 0$ then $m'_{i,j} = \bot$ and if $F(k, i, j) = 1$ then $m'_{i,j} = m_{i,j}$.
- Finally, for each agent i, $s'_i = \delta_i(s_i, P_i(s_i), (m'_{1,i}, \ldots, m'_{n,i}))$.

The system $\mathcal{R}_{\mathcal{E},\mathcal{F},P}$ consists of all runs generated from some initial state.

4 KNOWLEDGE-BASED PROGRAMS

Knowledge-based programs specify how an agent's actions are determined, given what the agent knows. As defined by Fagin et al. [5], these programs are interpreted relative to an *interpreted context* that defines the global states, how they are updated as a result of actions, and an interpretation of atomic propositions. In our setting, we can take the interpreted context to be a tuple $(\mathcal{E}, \mathcal{F}, \pi)$ consisting of an information-exchange protocol $\mathcal{E}$, a failure model $\mathcal{F}$, and an interpretation π of atomic propositions in the set of all runs over global states constructed from $\mathcal{E}$ and $\mathcal{F}$.

For our purposes, it is convenient to take knowledge-based programs to have the form $\mathbf{P} = (\mathbf{P}_1, \ldots, \mathbf{P}_n)$, where for each agent i, the local knowledge-based program $\mathbf{P}_i$ is in the language with grammar

$$\mathbf{P}_i ::= a_i \mid \text{if } \phi_i \text{ then } \mathbf{P}_i \text{ else } \mathbf{P}_i,$$

where a_i denotes actions in the set A_i of actions of agent i, and ϕ_i is a Boolean combination of formulas of the form $K_i\psi$. That is, the tests in agent i's local knowledge-based program concern agent i's knowledge. Note that the truth of such a formula ϕ_i at a point (r, m) in an interpreted system $\mathcal{I}$ depends only on agent i's local state at that point. That is, for points (r, m), $(r'm')$ with $r_i(m) = r'_i(m')$, we have $\mathcal{I}, (r, m) \models \phi_i$ iff $\mathcal{I}, (r', m') \models \phi_i$. Given a local state s of agent i, we may therefore write $\mathcal{I}, s \models \phi_i$ to express that $\mathcal{I}, (r, m) \models \phi_i$ for all points (r, m) of $\mathcal{I}$ with $r_i(m) = s$.

To interpret a knowledge-based program semantically, we first define how a knowledge-based program $\mathbf{P} = (\mathbf{P}_1, \ldots, \mathbf{P}_n)$ determines a concrete action protocol $\mathbf{P}^{\mathcal{I}}$ given an interpreted system $\mathcal{I}$. For each agent i and local state s of i in $\mathcal{I}$, we define $\mathbf{P}_i^{\mathcal{I}}(s)$ to be the action resulting from executing the program $\mathbf{P}_i$ with its tests interpreted at local state s in $\mathcal{I}$. Formally, we define $\mathbf{P}^{\mathcal{I}}$ by induction on the structure of $\mathbf{P}$, taking $(a_i)^{\mathcal{I}}(s) = a_i$, and for $\mathbf{P}_i$ ="if ϕ_i then $\mathbf{Q}_i$ else $\mathbf{R}_i$", we define $\mathbf{P}_i^{\mathcal{I}}(s) = \mathbf{Q}_i^{\mathcal{I}}(s)$ if $\mathcal{I}, s \models \phi_i$, and $\mathbf{P}_i^{\mathcal{I}}(s) = \mathbf{R}_i^{\mathcal{I}}(s)$ otherwise.

Given a knowledge-based program $\mathbf{P} = (\mathbf{P}_1, \dots, \mathbf{P}_n)$ and a concrete action protocol $P = (P_1, \dots, P_n)$ for an information-exchange protocol $\mathcal{E}$, we say that P *implements* $\mathbf{P}$ in the context $\gamma = (\mathcal{E}, \mathcal{F}, \pi)$ if, for $\mathcal{I} = (\mathcal{R}_{\mathcal{E},\mathcal{F},P}, \pi)$, we have $P_i(s) = \mathbf{P}_i^{\mathcal{I}}(s)$ for all agents $i = 1, \dots, n$ and local states s of agent i that arise in $\mathcal{I}$.

5 EVENTUAL BYZANTINE AGREEMENT

We briefly review the specification of the eventual Byzantine agreement problem that we consider in this paper. The specification assumes that each agent starts with an independently selected value $init_i \in \{0, 1\}$. The actions in A_i have the form $\mathtt{decide}_i(v)$, where $v \in \{0, 1\}$, as well as a "do-nothing" action noop. We seek protocols (i.e., an information-exchange protocol and an action protocol) such that every run satisfies the following four properties:

- **Unique Decision:** If agent i performs an action $\mathtt{decide}_i(v)$ (for some v), then it does not later perform $\mathtt{decide}_i(1 - v)$.
- **Agreement:** If agents i and j are both nonfaulty, i performs $\mathtt{decide}_i(v)$, and j performs $\mathtt{decide}_j(v')$, then $v = v'$.
- **Validity:** If a nonfaulty agent i performs $\mathtt{decide}_i(v)$ then $init_j = v$ for some agent j.
- **Termination:** For all nonfaulty agents i, eventually i performs $\mathtt{decide}_i(v)$ for some value $v \in \{0, 1\}$.

To relate this specification to our formal model, we define an *EBA context* to be a tuple $(\mathcal{E}, \mathcal{F}, \pi)$ consisting of an information-exchange protocol $\mathcal{E}$, a failure model $\mathcal{F}$, and an interpretation π of a set *Prop* of propositions, such that the following conditions hold:

- The local states in $\mathcal{E}_i$ have the form $\langle time_i, init_i, decided_i, jd_i, \dots \rangle$, where $time_i \in \mathbb{N}$, $init_i \in \{0, 1\}$, $decided_i \in \{0, 1, \bot\}$ and $jd_i \in \{0, 1, \bot\}$. Intuitively, $jd_i = v$ if i learned that some agent just decided v, for $v \in \{0, 1\}$.
- The initial local states in $\mathcal{E}_i$ have the form $\langle 0, init_i, \bot, jd_i^0, \dots \rangle$, where $jd_i = \bot$.
- The message-selection value $\mu_i(s, a)$ satisfies the following constraint: i sends different messages in the following three cases: (a) $a = \mathtt{decide}_i(0)$, (b) $a = \mathtt{decide}_i(1)$, and (c) the remaining cases. That means that j can tell from the message it receives from i whether i is deciding 0 or 1 in the current round. Formally, this means that there are three disjoint sets M^0, M^1, and M^2 with $\bot \notin M^0 \cup M^1$ such that if $a = \mathtt{decide}_i(0)$, then i sends each agent j a message in M^0, if $a = \mathtt{decide}_i(1)$, then i sends each agent j a message in M^1, and otherwise, i sends each agent j a message in M^2.
- The transition function δ_i, when given as input state s, action a, and a message tuple $(m_1, \dots, m_n)$, increases the time component $time_i$ of s by 1; if $a = \mathtt{decide}_i(v)$, it sets $decided_i$ to v, and otherwise leaves $decided_i$ unchanged; it also sets $jd_i = 0$ if i received a message in round m from an agent that performs action $\mathtt{decide}_i(0)$ in that round, sets $jd_i = 1$ if i received a message in round m from an agent that performs action $\mathtt{decide}_i(1)$ in that round, and otherwise sets $jd_i = \bot$ (the assumptions on μ_i ensure that such messages are distinguishable from other messages). Note that the fact that the *time* component increases by 1 at every step ensures that the system is *synchronous*; all agents i have $time_i = m$ at time m.
- *Prop* contains at least the following propositions (all of which are necessary to define the specification below) for each agent $i \in Agt$:
  - $init_i = v$ for $v \in \{0, 1\}$,
  - $decided_i = v$ for $v \in \{\bot, 0, 1\}$,
  - $time_i = k$ for $k \in \mathbb{N}$, and
  - $i \in \mathcal{N}$;

- π interprets $init_i = v$, $decided_i = v$, and $time_i = k$ in the obvious way from i's local state (e.g., $\pi(r, m)$ makes $init_i = v$ true iff $r_i(m)$ has its second component $init_i$ equal to v), and interprets $i \in \mathcal{N}$ in the obvious way from $\mathcal{F}$ (i.e., $\pi(r, m)$ makes $i \in \mathcal{N}$ true iff $i \in \mathcal{N}(r)$, where $\mathcal{N}(r)$ is the set of nonfaulty agents in run r).

An EBA context satisfies some minimal properties that we expect all contexts that arise in the analysis of EBA to satisfy.

We define $decided_i$ to be an abbreviation of $decided_i = 0 \lor decided_i = 1$, take $jdecided_i = v$ to be an abbreviation for $decided_i = v \land \ominus decided_i = \bot$ (intuitively, i just decided v), take $deciding_i = v$ to be an abbreviation for $decided_i = \bot \land \bigcirc decided_i = v$ (intuitively, i is deciding v in the current round) and take $\exists v$, for $v \in \{0, 1\}$ to be an abbreviation of $\bigvee_{i \in Agt} init_i = v$.

Given an EBA context $\gamma = (\mathcal{E}, \mathcal{F}, \pi)$ and an action protocol P for $\mathcal{E}$, we define the system $\mathcal{I}_{\gamma,P} = (\mathcal{R}_{\mathcal{E},\mathcal{F},P}, \pi)$. To satisfy the informal specification above, we now seek an EBA-context $\gamma = (\mathcal{E}, \mathcal{F}, \pi)$ and an action protocol P for $\mathcal{E}$ such that the following are valid in the system $\mathcal{I}_{\gamma,P}$ for all agents i and j:

- **Unique Decision:** $decided_i = v \Rightarrow \Box\neg(decided_i = 1 - v)$, for $v \in \{0, 1\}$.
- **Agreement:** $\neg(i \in \mathcal{N} \land j \in \mathcal{N} \land decided_i = v \land decided_j = v')$, for $v \neq v'$.
- **Validity:** $(decided_i = v \land i \in \mathcal{N}) \Rightarrow \exists v$
- **Termination:** $i \in \mathcal{N} \Rightarrow \Diamond(decided_i \neq \bot)$.

If these conditions are satisfied, we call P an *EBA decision protocol* for the context γ. The tuple $(\mathcal{E}, P, \pi)$ is an *EBA-protocol* for failure model $\mathcal{F}$. That is, a protocol solving EBA in the failure model consists of an information-exchange protocol, an action protocol that makes decisions, and an interpretation of the basic propositions.

We are interested in protocols that are optimal given the information that is maintained by the information-exchange protocol. The following definitions formalize this notion. Runs r, r' of two action protocols P, P', respectively, *correspond* if $r(0) = r'(0)$. That is, the two runs have the same failure pattern and the same initial states for all agents. Recall that the initial global state of a run, the information-exchange, and the action protocol together determine the complete run. An action protocol P *dominates* action protocol P' with respect to a context $\gamma = (\mathcal{E}, \mathcal{F}, \pi)$, written $P' \leq_\gamma P$ if, for all corresponding runs $r \in \mathcal{R}_{\mathcal{E},\mathcal{F},P}$ and $r' \in \mathcal{R}_{\mathcal{E},\mathcal{F},P'}$ and all agents i that are nonfaulty in r and times m, if $P_i(r_i(m)) = \mathtt{decide}_i(v)$ for $v \in \{0, 1\}$ then $P'_i(r'_i(m')) \neq \mathtt{decide}_i(w)$ for any $m' < m$ and $w \in \{0, 1\}$. That is, P makes it decisions no later than P'. P *strictly dominates* P' with respect to γ if $P' \leq_\gamma P$ and it is not the case that that $P \leq_\gamma P'$. An EBA decision protocol P is *optimal* with respect to an EBA context γ if no EBA decision protocol for γ strictly dominates P.

6 OPTIMAL EBA WITH RESPECT TO LIMITED INFORMATION EXCHANGE

In this section, we describe a knowledge-based program for EBA that is somewhat biased towards 0, show that it is correct, and show that it is optimal with respect to all information-exchange protocols that satisfy a certain safety condition. As discussed in the introduction, there is no protocol for EBA in the presence of omission failures where an agent decides 0 as soon as it hears that some agent had an initial preference of 0. So we consider instead a program where an agent decides 0 if it hears that some agent had an initial preference of 0 via a chain of agents; this is essentially the condition used to decide 0 in the crash-failure case.

A sequence $i_0, \ldots, i_m$ of distinct agents is a *0-chain of length m in run r of interpreted system $\mathcal{I}$* if (a) $\mathcal{I}, (r, 0) \models init_{i_0} = 0$, (b) for all m' with $0 \leq m' \leq m$, agent $i_{m'}$ first decides 0 in round $m' + 1$ of r, and (c) for all m' with $1 \leq m' \leq m$, $i_{m'}$ knows at the point (r, m') that $i_{m'-1}$ has just decided 0. We say that *an agent i receives a 0-chain in round m* in run r if there is a 0-chain of length m that ends with agent i in run r.

Variants of what we call a 0-chain exist in the literature [1, 8]. The 0-chain definition in [1], defined for crash failures, requires only that $i_{m'}$ receives a message from $i_{m'-1}$ in round m' for $m' \geq 1$. The 0-chain definition in [8] requires that i_m receives a message from i_{m-1} in round m' for $m' \geq 1$ and i_m considers it possible that i_{m-1} is nonfaulty when it receives the message (this will automatically be the case with crash failures, which is all that are considered in [1], but is not necessarily the case with omission failures, which are considered in [8]); moreover, i_m is required to be nonfaulty.

Let $\mathbf{P}_i^0$ be the following knowledge-based program for agent i:

Program: $\mathbf{P}_i^0$

```
if $decided_i \neq \bot$ then noop
else if $init_i = 0 \lor K_i(\bigvee_{j \in Agt} jdecided_j = 0)$ then $\mathtt{decide}_i(0)$
else if $K_i(\bigwedge_{j \in Agt} \neg(deciding_j = 0))$ then $\mathtt{decide}_i(1)$
else noop
```

In words: as long as i hasn't already decided, then i decides 0 if it has an initial preference of 0 or knows that someone just decided 0; i decides 1 if it knows that no agent can be currently deciding 0; otherwise, it does nothing. $\mathbf{P}^0$ is essentially the same as the knowledge-based program used by Castañeda et al. [1] in the case of crash failures. We will show that the second condition for deciding 0, that i knows that someone has just decided 0, holds iff i receives a 0-chain. It follows that if i hasn't already decided and is not deciding 0, then i decides 1 iff i knows that no agent is receiving a 0-chain. This latter condition is very close in spirit to Castanñeda et al.'s notion of there being no *hidden paths*.

$\mathbf{P}^0$ satisfies all the EBA properties in all EBA contexts.

Proposition 6.1. *If $\gamma = (\mathcal{E}, \mathcal{F}, \pi)$ is an EBA context, then all implementations of the knowledge-based program $\mathbf{P}^0$ with respect to γ are EBA decision protocols for γ. Indeed, all implementations of $\mathbf{P}^0$ terminate after at most $t + 1$ rounds of message exchange and Validity holds even for faulty agents.*

We next prove that some implementations of $\mathbf{P}^0$ are actually optimal EBA decision protocols in certain EBA contexts. Instead of individually proving optimality of $\mathbf{P}^0$ with respect to specific information exchanges, we first give a sufficient condition for optimality in EBA contexts. We then look at two specific contexts that satisfy this safety property and show that implementations of $\mathbf{P}^0$ in them are optimal.

Definition 6.2 (safety). A knowledge-based program $\mathbf{P}$ is *safe with respect to an EBA context $\gamma = (\mathcal{E}, \mathcal{F}, \pi)$* if, for all implementations P of $\mathbf{P}$ with respect to γ and all points (r, m) of $\mathcal{I} = (\mathcal{R}_{\mathcal{E},\mathcal{F},P}, \pi)$, the following two conditions hold:

(1) If i has not received a 0-chain by (r, m), then there exists a point (r', m) such that $r_i(m) = r'_i(m)$ and all agents have initial preference 1 in r'.

(2) If $\mathcal{I}, (r, m) \models \neg K_i(\bigwedge_{j \in Agt} \neg(deciding_j = 0))$ and i does not decide before round $m + 1$ in r, then there exists a point (r', m) such that:

  (a) $r_i(m) = r'_i(m)$,

  (b) i is nonfaulty in r',

  (c) some agent j that is nonfaulty in r' decides 0 in round $m+1$ of r'; moreover, if $m \geq 1$, there exists a run r'' and an agent j' such that j and j' are nonfaulty in r'', $r'_j(m) = r''_j(m)$, and j' decides 0 in round m in round r''.

Intuitively, (1) says that the only way that an agent learns that some agent had an initial preference of 0 is via a 0-chain; thus, (1) implies that if i has not received a 0-chain by (r, m), then $I, (r, m) \models \neg K_i \exists 0$. Clause (2) says that the only way an agent is unable to decide 1 (i.e., the test for deciding 1 in $\mathbf{P}^0$ does not hold) is if it considers it possible that some nonfaulty agent is deciding 0; thus, (2) implies that if $\mathcal{I}, (r, m) \models \neg K_i(\bigwedge_{j \in Agt} \neg(deciding_j = 0))$ and i has not decided by round m, then $\mathcal{I}, (r, m) \models \neg K_i(\bigwedge_{j \in \mathcal{N}} \neg(deciding_j = 0))$.

Note that a knowledge-based program will in general not be safe with respect to an FIP, since an agent i may learn that some agent j had an initial preference of 0 without receiving a 0-chain. As a result, the first condition will not hold, since at all points that i considers possible, j has an initial preference of 0. (In Section 7, we show that a small modification of $\mathbf{P}^0$ is optimal even with full information exchange.) But, as we shall see, $\mathbf{P}^0$ is safe with respect to two EBA contexts of interest, where agents do not keep track of who is faulty.

Theorem 6.3. *If γ is an EBA context and $\mathbf{P}^0$ is safe with respect to γ, then all implementations of $\mathbf{P}^0$ are optimal with respect to γ.*

We now describe two families of concrete EBA contexts with respect to which $\mathbf{P}^0$ is safe, parameterized by the number n of agents involved. For the first, let $\mathcal{E}_{\min}(n)$ be the *minimal* information-exchange protocol for n agents, where for each agent i, the following hold:

- The local states have the form $\langle time_i, init_i, decided_i, jd_i \rangle$. Thus, the local states include just what is required in an EBA context
- The initial local states have the form $\langle 0, init_i, \bot, \bot \rangle$.
- $M_i = \{0, 1\}$, $M^0 = \{0\}$, $M^1 = \{1\}$ and $M^2 = \{\bot\}$.
- For each agent j, if $a = \texttt{decide}_i(v)$ then $\mu_{ij}(s, a) = v$; otherwise, $\mu_{ij}(s, a) = \bot$. Note μ_{ij} satisfies the constraint we imposed for EBA contexts. Intuitively, if $\mu_{ij}(s, a) = v \neq \bot$, then i is about to decide v.
- The state-update component δ_i is defined on $time_i$, $init_i$, $decided_i$, and jd_i just as in EBA contexts.

For the second, let $\mathcal{E}_{basic}(n)$ be the *basic* information-exchange protocol for n agents, where the local states of agents are like those in a minimal information-exchange protocol, except that, in addition to the other messages allowed in a basic information-exchange protocol, each agent i can send a message of the form $(init, 1)$, and their local states have one additional component, $\#1_i$, that intuitively counts how many messages of the form $(init, 1)$ i received in the last round. In more detail, for each agent i, the following hold:

- The local states have the form $\langle time_i, init_i, decided_i, jd_i, \#1_i \rangle$, where $\#1_i \in \{0, \ldots, n\}$.
- The initial local states have the form $\langle 0, init_i, \bot, \bot, 0 \rangle$.
- $M_i = \{0, 1, (init, 1)\}$, $M^0 = \{0\}$, $M^1 = \{1\}$, and $M^2 = \{(init, 1), \bot\}$.
- For all agents j, if $a = \texttt{decide}_i(v)$ then $\mu_{ij}(s, a) = v$ for $v \in \{0, 1\}$; if $a = \texttt{noop}$ and s has the form $\langle m, 1, \bot, \bot, k \rangle$, then $\mu_{ij}(s, a) = (init, 1)$; otherwise, $\mu_{ij}(s, a) = \bot$.
- The state-update component δ_i is defined as in EBA contexts, with the added constraint that $\#1_i$ is updated to the number of messages of the form $(init, 1)$ that i receives in the current round if $decided_i = \bot$ and i does not receive a message $v \in \{0, 1\}$ from some agent j; otherwise, $\#1_i$ is set to 0 (essentially, for technical reasons, once a decision is made, $\#1_i$ is ignored).

Let $\gamma_{\min,n,t} = (\mathcal{E}_{\min}(n), SO(t), \pi_{\min,n})$ denote the family of minimal contexts where there are n agents, at most $t < n$ faulty agents, the language includes $jd_i = v$ in addition to $time_i = k$, $init_i = v$, $decided_i(v)$, and $i \in \mathcal{N}$, for $i \in \{1, \ldots, n\}$, and $\pi_{\min,n}$ interprets these primitive propositions in the obvious way. Similarly, let $\gamma_{basic,n,t} = (\mathcal{E}_{basic}(n), SO(t), \pi_{basic,n})$ denote the family of basic contexts with n agents and $t < n$ faulty agents, where the language includes $\#1_i = k$ in addition to all the primitive proposition used in minimal contexts, where again, $\pi_{basic,n}$ interprets all the primitive propositions in the obvious way.

Proposition 6.4. *$\mathbf{P}^0$ is safe with respect to all contexts $\gamma_{\min,n,t}$ and $\gamma_{basic,n,t}$ such that $n - t \geq 2$.*

Finally, we provide EBA decision protocols that implement $\mathbf{P}^0$ in the two contexts of interest. Let P^{min} be the protocol implemented by the following (standard) program:

Program: P_i^{min}

```
if $decided_i \neq \bot$ then noop
else if $init_i = 0 \vee jd_i = 0$ then decide$_i(0)$
else if $time_i = t + 1$ then decide$_i(1)$
else noop
```

Intuitively, this EBA decision protocol decides 0 if the agent has initial value 0 or hears of a 0-decision by another agent. If the agent does not hear about a 0-decision by time $t + 1$, then it decides 1.

Theorem 6.5. *If $t \leq n - 2$, then P^{min} implements $\mathbf{P}^0$ in the EBA context $\gamma_{\min,n,t}$.*

Finally, let P^{basic} be the EBA decision protocol that implements the following program:

Program: P_i^{basic}

```
if $decided_i \neq \bot$ then noop
else if $init_i = 0 \vee jd_i = 0$ then decide$_i(0)$
else if $\#1_i > n - time_i \vee jd_i = 1$ then decide$_i(1)$
else noop
```

Theorem 6.6. *If $t \leq n - 2$, then P^{basic} implements $\mathbf{P}^0$ in the EBA context $\gamma_{basic,n,t}$.*

We get the following immediate corollary to Theorems 6.3, 6.5, and 6.6.

Corollary 6.7. *P^{basic} is optimal with respect to $\gamma_{basic,n,t}$ and P^{min} is optimal with respect to $\gamma_{\min,n,t}$.*

7 AN OPTIMAL FULL-INFORMATION PROTOCOL FOR EBA

While $\mathbf{P}^0$ is optimal with respect to the basic and minimal information-exchange settings, it is not optimal in the full-information setting, as the following example shows.

Example 7.1. Suppose that $t = 10$ and $n = 20$. Consider a run where all agents have initial value 1, agents 1–10 are faulty, and no faulty agent sends a message in any round. This means that, at the end of the first round, each nonfaulty agent knows who the faulty agents are. At the end of the second round, it is common knowledge among the nonfaulty agents who the faulty agents are: each nonfaulty agent i will know at that point that each other nonfaulty agent j knows who the nonfaulty agents are and (by the same reasoning) i knows that j knows that all the nonfaulty agents know who the faulty agents are, and so on. Moreover, it is common knowedge among the nonfaulty agents that no nonfaulty agent has already decided, and it is not common knowledge among the nonfaulty agents that some agent had an initial preference of 0, while it is common knowledge that some agent had an initial preference of 1. As we show below, this means that the nonfaulty agents can decide on 1 in round 3. On the other hand, with P^{basic} and P^{min}, the nonfaulty agents would not decide in this run until round 12.

Intuitively, we take $C_{\mathcal{N}}(\textit{t-faulty})$ to hold when it is common knowledge among the nonfaulty agents who the nonfaulty agents are. It turns out this can happen only if the nonfaulty agents have common knowledge of t agents that are faulty. Thus, we take $C_{\mathcal{N}}(\textit{t-faulty} \wedge \phi)$, for each formula ϕ, to be an abbreviation for $\exists A \subseteq Agt(|A| = t \wedge C_{\mathcal{N}}(\bigwedge_{i \in A}(i \notin \mathcal{N}) \wedge \phi))$. We also define $\textit{no-decided}_{\mathcal{N}}(x)$ as an abbreviation for $\bigwedge_{j \in \mathcal{N}} \neg(\textit{decided}_j = x)$ for $x \in \{0,1\}$. We can now formalize the situation in the example.

Proposition 7.2. *If P is an optimal protocol in the context $\gamma_{\mathit{fip},n,t}$ and $\mathcal{I}_{P,\gamma_{\mathit{fip},n,t}}, (r,m) \models \textit{decided}_i = \bot \wedge K_i(C_{\mathcal{N}}(\textit{t-faulty} \wedge \textit{no-decided}_{\mathcal{N}}(1) \wedge \exists 0))$, then all undecided agents in $\mathcal{N}(r)$ make a decision in round $m+1$, and similarly if $\mathcal{I}_{P,\gamma}, (r,m) \models \textit{decided}_i = \bot \wedge K_i(C_{\mathcal{N}}(\textit{t-faulty} \wedge \textit{no-decided}_{\mathcal{N}}(0) \wedge \exists 1))$.*

It turns out that if we add a condition to $\mathbf{P}^0$ that tests for this common knowledge and decides appropriately if it holds, we get a program that is optimal even with full information exchange. Specifically, let $\mathbf{P}^1_i$ be the following knowledge-based program for agent i:

Program: $\mathbf{P}^1_i$

```
if $decided_i \neq \bot$ then noop
else if $K_i(C_{\mathcal{N}}(\textit{t-faulty} \wedge \textit{no-decided}_{\mathcal{N}}(1) \wedge \exists 0))$ then $\texttt{decide}_i(0)$
else if $K_i(C_{\mathcal{N}}(\textit{t-faulty} \wedge \textit{no-decided}_{\mathcal{N}}(0) \wedge \exists 1))$ then $\texttt{decide}_i(1)$
else if $init_i = 0 \vee K_i(\bigvee_{j \in Agt} jdecided_j = 0)$ then $\texttt{decide}_i(0)$
else if $K_i(\bigwedge_{j \in Agt} \neg(deciding_j = 0))$ then $\texttt{decide}_i(1)$
else noop
```

Note that in the basic and mininimal contexts, agents never learn who is faulty, so there is never common knowledge among the nonfaulty agents who the faulty agents are. Thus, in the contexts $\gamma_{\min,n,t}$ and $\gamma_{basic,n,t}$, $\mathbf{P}^1$ is equivalent to $\mathbf{P}^0$, so $\mathbf{P}^1$ is correct and optimal in these contexts. As we are about to show, $\mathbf{P}^1$ is also correct and optimal with full information exchange.

To prove correctness and optimality, we follow the approach of Halpern, Moses, and Waarts [8] and consider a slightly nonstandard full-information context. We assume that each agent i's local state does *not* contain the variables $\textit{decided}_i$ and jd_i, but does contain a variable or variables that keep track of all messages received from all agents. If agents keep track of all messages received in their local state, then, given a decision protocol P, the variables $\textit{decided}_i$ and jd_i are redundant; their values can be inferred from the messages received. Let $\gamma_{\mathit{fip},n,t}$ denote the family of full-information contexts as described above. Not including $\textit{decided}_i$ and jd_i in the local state has the advantage that, for all decision protocols P and P', corresponding runs of P and P' in $\gamma_{\mathit{fip},n,t}$ are actually identical; although agents may make different decisions, their local states are the same at all times. (This would not be the case if the local states had included information about decisions, and in particular, if they had included the variables $\textit{decided}_i$ and jd_i.) It is critical that we are dealing with FIPs here; the claim is not true for arbitrary information-exchange protocols.

Proposition 7.3. *All implementations of $\mathbf{P}^1$ with respect to $\gamma_{\mathit{fip},n,t}$ are EBA decision protocols for $\gamma_{\mathit{fip},n,t}$.*

We need to recall some material from [8] in order to use the characterization of optimality with respect to $\gamma_{\mathit{fip},n,t}$. Given an *indexical set* $\mathcal{S}$ of agents, a point (r', m') is $\mathcal{S}$-$\boxdot$-*reachable from* (r, m)

if there exist runs $r^0, \dots, r^k$, times $m_0, m_0', \dots, m_k, m_k'$, and agents $i_0, \dots, i_{k-1}$ such that $(r^0, m_0) = (r, m)$, $(r^k, m_k') = (r', m')$, and for $0 \le j \le k-1$, we have $i_j \in \mathcal{S}(r^j, m_j') \cap \mathcal{S}(r^{j+1}, m_{j+1})$ and $r_j^j(m_j') = r_j^{j+1}(m_{j+1})$.[2] Using the notation of [8], let $\mathcal{N} \wedge \mathcal{O}$ denote the indexical set where $(\mathcal{N} \wedge \mathcal{O})(r, m)$ consists of all agents that are nonfaulty and about to decide 1 or have already decided 1 at the point (r, m). Let $(\mathcal{N} \wedge \mathcal{Z})$ be the analogous set for 0.

Definition 7.4 (weak safety). A knowledge-based protocol $\mathbf{P}$ is *weakly safe with respect to an EBA context* γ if, for all implementations P of $\mathbf{P}$ and all points (r, m) of $\mathcal{I} = (\mathcal{R}_{\mathcal{E},\mathcal{F},P}, \pi)$ and all agents i, if $\mathcal{I}_{\gamma_{fip,n,t},P}, (r, m) \models i \in \mathcal{N} \wedge \bigcirc(decided_i = \bot)$, then there exist points $(r^{0\prime}, m)$, $(r^{0\prime\prime}, m)$, $(r^{1\prime}, m)$, and $(r^{1\prime\prime}, m)$ such that:

1. $r_i(m) = r_i^{0\prime}(m) = r_i^{1\prime}(m)$,
2. i is nonfaulty in $r^{0\prime}$ and $r^{1\prime}$,
3. $(r^{0\prime\prime}, m'')$ is $(\mathcal{N} \wedge \mathcal{Z})$-$\boxdot$-reachable from $(r^{0\prime}, m)$,
4. $(r^{1\prime\prime}, m'')$ is $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$-reachable from $(r^{1\prime}, m)$,
5. all agents have initial preference 0 in $r^{0\prime\prime}$,
6. all agents have initial preference 1 in $r^{1\prime\prime}$.

Our interest in weak safety is motivated by the following result proved by Halpern, Moses, and Waarts [8]. The statement of the result uses two operators, $B_i^{\mathcal{N}}$ $C_{\mathcal{S}}^{\boxdot}$. $B_i^{\mathcal{N}}$ is an abbreviation of $K_i(i \in \mathcal{N} \Rightarrow \phi)$. Thus, $\mathcal{I}, (r, m) \models B_i^{\mathcal{N}}\phi$ if and only if $\mathcal{I}, (r', m') \models \phi$ for all points (r', m') such that $r_i(m) = r_i'(m')$ and $i \in \mathcal{N}(r')$. Intuitively, $B_i^{\mathcal{N}}$ holds if i knows that if it is nonfaulty, then ϕ holds. The $C_{\mathcal{S}}^{\boxdot}$. operator has a characterization in terms of $\mathcal{S}$-$\boxdot$-reachability. In [8], it is shown that $\mathcal{I}, (r, m) \models C_{\mathcal{S}}^{\boxdot}\phi$ if and only if $\mathcal{I}, (r', m') \models \phi$ for all points (r', m') that are $\mathcal{S}$-$\boxdot$-reachable from (r, m). We are interested in the cases that $\mathcal{S}$ is either $\mathcal{N} \wedge \mathcal{O}$ or $\mathcal{N} \wedge \mathcal{Z}$.

Theorem 7.5. *[8, Theorem 5.4] An EBA protocol P is optimal with respect to $\gamma_{fip,n,t}$ iff the following two conditions hold:*

$$\mathcal{I}_{\gamma_{fip,n,t},P} \models i \in \mathcal{N} \Rightarrow \ (\bigcirc(decided_i = 0) \Leftrightarrow B_i^{\mathcal{N}}(\exists 0 \wedge C^{\boxdot}_{\mathcal{N}\wedge\mathcal{O}}\exists 0 \wedge \neg(\bigcirc(decided_i = 1))))$$
$$\mathcal{I}_{\gamma_{fip,n,t},P} \models i \in \mathcal{N} \Rightarrow \ (\bigcirc(decided_i = 1) \Leftrightarrow B_i^{\mathcal{N}}(\exists 1 \wedge C^{\boxdot}_{\mathcal{N}\wedge\mathcal{Z}}\exists 1 \wedge \neg(\bigcirc(decided_i = 0)))).$$

Using Theorem 7.5, we can prove that weak safety implies optimality for $\mathbf{P}^1$.

Theorem 7.6. *If $\mathbf{P}^1$ is weakly safe with respect to $\gamma_{fip,n,t}$ then all implementations of $\mathbf{P}^1$ are optimal with respect to $\gamma_{fip,n,t}$.*

To show that the knowledge-based program $\mathbf{P}^1$ satisfies weak safety with respect to the full-information context, we give a constructive proof that explicitly constructs the sequences of points witnessing the conditions of the definition of weak safety.

The main idea of the proof comes from the following observation. If a nonfaulty agent i is unable to decide, then the common knowledge conditions in $\mathbf{P}^1$ do not hold. Then, roughly speaking, we can show that there exists points $(r^{0\prime}, m)$ and $(r^{1\prime}, m)$, as in the definition of weak safety, such that a faulty agent k acts nonfaulty throughout the run.[3] Moreover, since this faulty agent k did not

[2] Halpern, Moses, and Waarts [8] introduced a family of *continual common knowledge* operators $C_{\mathcal{S}}^{\boxdot}$ such that $C_{\mathcal{S}}^{\boxdot}\phi$ holds at a point (r, m) iff ϕ is true at all points (r', m') that are $\mathcal{S}$-$\boxdot$-reachable from (r, m). We get standard (indexical) common knowledge by taking $m_k = m_k'$ in the definition of continual common knowledge; since we are working with synchronous systems, we could restrict to taking $m_j' = m_{j+1}$.

[3] Note that since the set of faulty agents in a run is determined by the failure pattern, it is consistent that an agent i is faulty in a run although it acts nonfaulty throughout the run. Since all that really matters for our result is that no agent can detect that agent i is faulty, we could obtain our result by assuming that i's faulty behavior involved only not sending messages to itself.

display faulty behavior, all runs where another agent is faulty instead of k are indistinguishable from this point. The existence of a faulty agent that acts nonfaulty turns out to be a strong condition that allows the construction of a sequence of $\boxdot$-reachable points where the end point has a modified message pattern in addition to satisfying the same condition on k. This is possible by temporarily making k exhibit faulty behavior in intermediate points of the sequence. Therefore we use the existence of such k as an invariant that allows taking steps through the $\boxdot$-path.

Theorem 7.7. *$\mathbf{P}^1$ is weakly safe with respect to $\gamma_{fip,n,t}$.*

Proposition 7.3, Theorem 7.6, and Theorem 7.7 together imply the following corollary.

Corollary 7.8. *All implementations of $\mathbf{P}^1$ with respect to $\gamma_{fip,n,t}$ are optimal with respect to $\gamma_{fip,n,t}$.*

Implementing $\mathbf{P}^1$ in polynomial time is possible using the compact communication graph representation of the full-information exchange due to [12]. Intuitively, the common knowledge conditions are implemented using the observation that if an agent's faultiness is common knowledge among the nonfaulty agents, it must be *distributed knowledge* at the previous time (where a fact ϕ is distributed knowledge among an indexical set $\mathcal{S}$ of agents if the agents would know ϕ if they pooled their knowledge together; for example, the set of faulty agents is distributed knowledge among the nonfaulty agents if, between them, the nonfaulty agents know who the faulty agents are). Since nonfaulty agents send messages describing their complete state in every round, we can check whether $C_{\mathcal{N}}(t\text{-}faulty)$ holds at a point (r, m) by considering the local states at $(r, m-1)$ of the agents that nonfaulty agents heard from in round m.

Proposition 7.9. *There exists a polynomial-time implementation P^{opt} of $\mathbf{P}^1$ with respect to a full-information exchange.*

8 DISCUSSION

We introduced the notion of limited information exchange, examined optimality for EBA with respect to various information-exchange protocols and described an efficiently implementable optimal FIP. There is clearly far more to be done. There are two short-term directions we are currently pursuing. First, we hope to explore the impact of limited information exchange on other protocols of interest. Second, we are exploring the application of epistemic synthesis techniques that allow the automated derivation of protocols from a knowledge-based program in the context of limited information-exchange models. This seems to give the techniques far more scope (cf. [9, 10]).

We conclude this discussion by taking a closer look at the costs and benefits of limited information exchange for EBA. We focus on the two settings considered in Section 6, as well as the full-information context, and consider the cost both in terms of the number of bits sent and the number of rounds required to reach a decision in the most likely case, where there are no failures. While the results are straightforward, they help highlight the tradeoffs involved. Let $\gamma_{fip,n,t}$ be a full-information EBA context with omission failures and let P^{fip} be an implementation of $\mathbf{P}^1$ in $\gamma_{fip,n,t}$.

We start by considering message complexity in terms of bits. In the minimal information-exchange protocol P^{min}, each message can be represented using a single bit and agents send a message only when they first decide, otherwise staying silent. Since each agent sends exactly one message in each run, and sends it to all the other agents, n^2 bits are sent altogether. In the basic information-exchange protocol P^{basic}, we still require only a constant number of bits to represent messages and agents send messages to all other agents as long as they are undecided, which means for at most $t + 1$ rounds. We then get the following result:

Proposition 8.1. *In each run of P^{min}, n^2 bits are sent in total; in each run of P^{basic}, at most $O(n^2 t)$ bits are sent in total.*

By way of contrast, a standard communication graph implementation of a full-information exchange uses $O(n^4 t^2)$ bits [12].

We next consider decision times. We focus on the failure-free case as, in most applications, the most common runs are runs with no failures.

Proposition 8.2. *If r is a failure-free run, then*

(a) If there is at least one agent with an initial preference of 0 in r, then all agents decide by round 2 with P^{min}, P^{basic}, and P^{fip}.

(b) If all agents have an initial preference of 1, then all agents decide by round $t+2$ with P^{min} and by round 2 with P^{basic} and P^{fip}.

Thus, for failure-free runs, the agents in the basic context and the full-information context decide at the same time. The only failure-free run where the basic context results in an earlier decision than the minimal context is the run where every agent starts with an initial preference of 1. As a result, implementing full-information and incurring a quadratic overhead in the number of bits never leads to an improvement for failure-free runs. If we assume that each configuration of initial preferences is equally likely, using the basic context over the minimal context for failure-free runs is only an improvement $1/2^n$ of the time.

If failure-free runs are sufficiently common, this suggests that the gain of using an FIP may not be worth the cost; even the tradeoff between P^{basic} and P^{min} is not so clear. We conjecture that even in runs with failures, P^{basic} may not be much worse than P^{fip}. This emphasizes the advantages of considering limited information exchange, and further motivates considering optimal protocols with limited information exchange more broadly.

A PROOFS

A.1 Proofs for Section 6

Proposition 6.1. *If $\gamma = (\mathcal{E}, \mathcal{F}, \pi)$ is an EBA context, then all implementations of the knowledge-based program $\mathbf{P}^0$ with respect to γ are EBA decision protocols for γ. Indeed, all implementations of $\mathbf{P}^0$ terminate after at most $t+1$ rounds of message exchange and Validity holds even for faulty agents.*

Proof. Fix an implementation P of $\mathbf{P}^0$ in $\gamma_{fip,n,t}$.

Unique-Decision follows from the fact that P_i makes at most one decision per round and the fact that whether a decision was made is recorded in the local state variable $decided_i$.

To see that Agreement holds, we first show by induction on m that if i has not decided before round $m+1$ and $\mathcal{I}, (r,m) \models init_i = 0 \lor K_i(\bigvee_{j \in Agt} jdecided_j = 0)$, then i receives a 0-chain in round m. If $m = 0$, then it must be the case that $\mathcal{I}, (r,m) \models init_i = 0$ which is a 0-chain with $i_1 = i$. If $m > 0$, we cannot have $\mathcal{I}, (r,m) \models init_i = 0$ (as i would have decided earlier) so we must have $\mathcal{I}, (r,m) \models K_i(jdecided_j = 0)$ for some $j \in Agt$. Then, $\mathcal{I}, (r,m) \models jdecided_j = 0$ and the result follows from the induction hypothesis.

Suppose by way of contradiction that r is a run where there exist nonfaulty agents i and j and a time m such that $\mathcal{I}, (r,m) \models decided_i = 0 \land decided_j = 1$. Suppose that j decides 1 in round $m_j + 1$ and i decides 0 in round $m_i + 1$, so that the decision conditions first hold at times m_i and m_j, respectively. If $m_j \leq m_i$, must have either $\mathcal{I}, (r, m_i) \models init_i = 0$ or $\mathcal{I}, (r, m_i) \models K_i(jdecided_k = 0)$ for some $k \in Agt$. Using our observation, we can conclude that i receives a 0-chain at time m_i, which implies that there exists an agent i' such that $\mathcal{I}, (r, m_j) \models deciding_{i'} = 0$. Hence, $\mathcal{I}, (r, m_j) \models \neg K_j(\neg(deciding_{i'} = 0))$, so j cannot decide 1 at m_j. If $m_j > m_i$, since i decides 0 in round $m_i + 1$, we must have $\mathcal{I}, (r, m_i) \models init_i = 0 \bigvee_{j \in Agt} K_i(jdecided_j = 0)$. We can again use our observation to conclude that i receives a 0-chain at time m_i in r. As i is nonfaulty, j must hear from i in round $m_i + 1$, so $\mathcal{I}, (r, m_j) \models K_j(jdecided_i = 0)$. It follows that agent j should decide 0 in this run, contradicting the assumption that j decides 1.

For Validity, observe that if i decides 0, using the previous observation, there must be a 0-chain, and hence an agent that had an initial preference of 0. If agent i decides 1, then i did not decide 0 in the first round and therefore we must have $init_i = 1$. Note that this argument holds even if i is faulty.

Finally, we prove Termination by showing that in all runs, all nonfaulty agents must decide by round $t + 2$ after at most $t + 1$ rounds of message exchange. To see this, we first show that if some agent decides 0 in run r, then all agents that decide 0 must do so by round $t + 2$ after at most $t + 1$ rounds of message exchange. For suppose by way of contradiction that some agent i decides 0 in round $m > t + 2$. Then there must be a 0-chain $i_1, \ldots, i_m$ with $i_m = i$. All the agents on the 0-chain are distinct. Since there are at most t faulty agents in a run, one of $i_1, \ldots, i_{t+1}$ must be nonfaulty, say i_j. But that means that all agents (including i) would have received a message from i_j in round $j \leq t + 1$ from which they could infer that i_j is about to decide 0 (by our assumption regarding μ in an EBA context), and they would all decide 0 in round $j + 1 < m$ if they have not done so yet. This gives the desired contradiction. It follows that if i has not decided 0 by round $t + 2$, then $\mathcal{I}, (r, t + 2) \models K_i(\bigwedge_{j \in Agt} \neg(deciding_j = 0))$, so i will decide 1 in round $t + 2$ if it has not already decided. □

Theorem 6.3. *If γ is an EBA context and $\mathbf{P}^0$ is safe with respect to γ, then all implementations of $\mathbf{P}^0$ are optimal with respect to γ.*

Proof. Suppose by way of contradiction that P is an implementation of $\mathbf{P}^0$ with respect to an EBA context γ and EBA-protocol P' strictly dominates P with respect to γ. If P' strictly dominates P, by definition, there exist corresponding runs r of $\mathcal{I}_{\gamma,P}$ and r' of $\mathcal{I}_{\gamma,P'}$, a nonfaulty agent i, and a round k such that P' decides in round k of r' and P does not decide before round $k + 1$ in r. Clearly, there also exist corresponding runs r and r', an agent i (possibly faulty) and a round k such that P' decides in round k of r', and P either does not decide before round $k + 1$ of r, or decides differently in round k of r. Let k be the earliest such round, and let r and r' be the corresponding runs of P and P', respectively, where in r', P' decides in round k and in r, P either does not decide by round k or reaches a different decision in round k than P'. (Note that the k we use here may be smaller than the smallest k such that P' decides in round k and P does not decide before round $k + 1$ in corresponding runs.) Since the same information-exchange protocol is used in both systems, all agents must have exactly the same state up to time $k - 1$ in corresponding runs of the two systems: either they have not decided yet, or they have decided and made the same decision, so they will send the same messages and undergo the same state transitions in corresponding runs of P and P' up to time $k - 1$. We now consider two cases.

Suppose that i decides 0 in round k of r'. Since i either does not decide at or before round k of r or decides 1 in this round, i did not receive a 0-chain by $(r, k - 1)$. Since $\mathbf{P}^0$ is safe with respect to γ, by the first part of the safety condition, it follows that there is a run r_1 of P such that $r_i(k - 1) = (r_1)_i(k - 1)$ and all agents have initial preference 1 in r_1. Let r_1' be the run of P' corresponding to r_1. As observed above, i must have the same state at $(r_1, k - 1)$ and $(r_1', k - 1)$. Since i also has the same state in $(r, k - 1)$ and $(r', k - 1)$, i has the same state in $(r', k - 1)$ and $(r_1', k - 1)$, so must decide 0 in round k of r_1'. We now get the desired contradiction by observing that the decision rule for deciding 0 in $\mathbf{P}^0$ requires that there exists an agent with an initial preference 0. Note that this is because, as shown in Proposition 6.1, Validity holds even for faulty agents.

Suppose that i decides 1 in round k of r', does not decide 1 in round k of r, and does not decide earlier than round k in r. Then we must have that $\mathcal{I}_{\gamma,P}, (r,k-1) \models \neg K_i(\bigwedge_{j\in Agt} \neg(deciding_j = 0))$: If i decides 0 in round k, then clearly i considers it possible that some agent is deciding 0; and if i does not decide 0 and does decide 1 (the only other possibility), then this formula must also hold. Since $\mathbf{P}^0$ is safe with respect to γ, it follows that there is a run r_1 of P such that $r_i(k-1) = (r_1)_i(k-1)$, i is nonfaulty in r_1, and there exists a nonfaulty agent j that decides 0 in round k of r_1. Again, let r_1' be the run of P' corresponding to r_1. As above, i has the same state in $(r', k-1)$ and $(r_1', k-1)$, so must decide 1 in round k of r_1'. Since P' dominates P, j must also decide in round k of r_1', and must decide 1 (since i is deciding 1).

If $k = 1$, since j decides 0 in r_1, j must have an initial value of 0 in both r_1 and r_1'. Let r_1'' be the run of P' where all agents have initial value 0 and are nonfaulty. Since $(r_1'')_j(0) = (r_1')_j(0)$, j must decide 1 with P' in run r_1'', giving us the desired contradiction. If $k > 1$, by the second half of condition 2(c) of the safety condition, there exists a run r_2 of P and an agent j' such that $(r_2)_j(k-1) = (r_1)_j(k-1)$, j and j' are nonfaulty in r_2, and j' decides 0 in round $k-1$ of r_2. Let r_2' be the run of P' corresponding to r_2. As observed above, $(r_2)_j(k-1) = (r_2)_j'(k-1)$. Thus, $(r_1)_j'(k-1) = (r_2)_j'(k-1)$, so j must decide 1 in round k of r_2'. Since j' is nonfaulty and decides in round $k-1$ of r_2, and P' dominates P, agent j' must also decide at or before round $k-1$ of r_2'. Since, by construction, round k is the earliest round that P and P' reach different decisions in corresonding runs, j' must decide 0 in round $k-1$ of r_2'. But this means that j and j' make different decisions in r_2', despite both being nonfaulty. This gives us the desired contradiction. □

PROPOSITION 6.4. *$\mathbf{P}^0$ is safe with respect to all contexts $\gamma_{\min,n,t}$ and $\gamma_{basic,n,t}$ such that $n-t \geq 2$.*

PROOF. We do the argument simultaneously for $\gamma_{\min,n,t}$ and and $\gamma_{basic,n,t}$. Let P be an implementation of $\mathbf{P}^0$ in $\gamma_{\min,n,t}$ (resp., $\gamma_{basic,n,t}$) and let r be a run in $\mathcal{I}_{\gamma_{\min,n,t},P}$ (resp., $\mathcal{I}_{\gamma_{basic,n,t},P}$). We first show that the first part of the safety condition holds. Suppose that i has not received a 0-chain by (r,m). We want to show that there exists a point (r',m) such that $r_i(m) = r_i'(m)$ and all agents have initial preference 1 in r'. Let r' be the run where all agents start with initial preference 1 and the adversary is the same as in r. An easy argument by induction on k, using the fact that the failure pattern is the same in r and r', shows that that for all agents j and times k, if j has not received a 0-chain by (r,k) then $r_j(k) = r_j'(k)$. It immediately follows that $r_i(m) = r_i'(m)$, which completes the proof of the first part of the argument.

To prove that the second part of the safety condition holds, suppose that $\mathcal{I}, (r,m) \models \neg(K_i(\bigwedge_{j\in Agt} \neg(deciding_j = 0))$ and i does not decide before round $m+1$ of r. The second part of the safety condition is easily seen to hold if $m = 0$, so we assume that $m \geq 1$. This implies that $init_i = 1$, otherwise agent i would have already decided in round 1. There must exist a point (r^+, m) such that $r_i(m) = r_i^+(m)$ and some agent j decides 0 in round $m+1$ of r^+. Since $n-t \geq 2$, there must be agents, say i' and j', that are nonfaulty in r^+, where we can take $i = i'$ if i is nonfaulty in r^+, and $j = j'$ if j is nonfaulty in r^+. If $m > 1$ then the initial preferences of i' and j' in r^+ are 1, for otherwise i would decide 0 at or before round 2 in r^+, and hence also in r, contradicting the assumption i does not decide before round $m+1$ in r. The initial preference of j must also be 1, for otherwise j decides in round $1 \neq m+1$, given that $m \geq 1$.

Observe that in r^+, (a) no agent decides 1 at or before round $m+1$ (for if agent j'' decides 1 in round k of r^+, then $\mathcal{I}, (r^+, k-1) \models K_{j''}(\bigvee_{j'''\in Agt} \neg(deciding_{j'''} = 0))$, and this contradicts the fact that j decides 0 in round $m+1$ of r, so there must be a 0-chain of length m ending with j in r), (b) no nonfaulty agent decides 0 before round m in r^+ (otherwise i and j would decide 0 at or before round m in r^+), (c) i and j do not decide 0 at or before round m in r^+, and (d) i, j, and all the nonfaulty agents send no message (i.e., send $\perp$) in $\gamma_{\min,n,t}$ and send $(init, 1)$ in $\gamma_{basic,n,t}$ up to and including round $m-1$; i and j also send $\perp$ (resp., $(init, 1)$) in round m of r in $\gamma_{\min}(n)$ (resp., $(\gamma_{basic,n,t})$. Note that i' or j' may decide 0 in round m of r^+, in which case they will send 0; otherwise, like i and j, they send $\perp$ (resp., $(init, 1)$) in round m of r in $\gamma_{\min,n,t}$ (resp., $(\gamma_{basic,n,t})$.

We want to modify r^+ to get a run r' such that (a) $r_i(m) = r_i'(m)$, (b) i and j are nonfaulty in r', and (c) j decides 0 in round $m+1$ of r'. Let $\alpha = (\mathcal{N}', F')$ be the adversary in r'. We define r' by assuming that all agents have the same initial preferences in r' as in r^+, and the adversary $(\mathcal{N}'', F'')$ in r' is defined as follows: $\mathcal{N}'' = \mathcal{N}' - \{i', j'\} \cup \{i, j\}$ (so that i and j are nonfaulty in r') and, roughly speaking, F'' interchange the failures of i and i' and of j and j' in r^+ and r'. More precisely, for all agents j'', (a) if j'' does not receive a message from i (resp., j) in round k of r^+ according to F', and, in the special case that $k = m$, neither i' nor j' sends the message 0, then j'' does not receive a message from i' (resp., j') in round k of r' according to F''; (b) if j'' receives a message from i (resp., j) in round of r^+ according to F' or if $k = m$ and either i' or j' send the message 0 in round m of r, then j'' receives a message from i' (resp., j') in round k of r' according to F''.

We claim that for all $k \leq m$, all agents j'' have the same state at time k of runs r^+ and r'. We prove this by induction on k. In the case that $k = 0$, this is immediate since all agents have the same initial preferences in r^+ and r'. If $0 < k < m$, this follows from the fact that all agents have the same state at time $k-1$, and the only way in which the runs differ is that j'' does not receive a message from i (resp., j) in round k of r^+, then j'' does not receive a message from i' (resp., j') in round k of r'. These message are either $\perp$ (if the context is $\gamma_{\min,n,t}$) or $(init, 1)$ (if the context is $\gamma_{basic,n,t}$). But clearly this difference does not affect the state of j''; in particular, if the context is $\gamma_{basic,n,t}$, then j'' gets the same number of messages of the form $(init, 1)$ in both cases, and this is all it keeps track of in its state. The same argument applies if $k = m$ and neither i' or j' send the message 0 in round m of r^+. If i' or j' do send the message 0 in round m of r^+, since they are nonfaulty in r^+, all agents will get the message. By construction, they will also send this message in round m of r' and all agents will get it in r'. The transition function then guarantees that all agents will have the same state in round m of r^+ and r'. In particular, it follows that $r_i^+(m) = r_i'(m)$ and j decides 0 in round $m+1$ of r', as desired.

Finally, if $m \geq 1$, we must construct a run r'' as required in the second part of condition 2(c). Since j decides 0 in r' in round $m+1$, as shown earlier, j must receive a 0-chain at time m in r. Let j' be the last agent on that chain. Thus, j' decides 0 in round m in r'. If j' is nonfaulty in r', we are done. If j' is faulty, then we consider two cases. If $m = 1$, then we must have $init_{j'} = 0$. We consider a run r'' where all agents have the same initial values as in r', and if (N', F') is the adversary in r', then the adversary in r'' is $(N' - \{j'\}, F'')$, where F'' agrees with F' on all the agents in $N' - \{j'\}$. It is easy to see that $r_j'(1) = r_j''(1)$, completing the argument. If $m > 1$, then there must be some nonfaulty agent j'' other than i, since $n - t \geq 2$. As in the argument above, the initial values of j' and j'' must be 1 (otherwise j' would have decided in round 1 and i would have decided in round 2 in r'). We now proceed much as in the previous argument to construct r'': all agents have the same initial values in r' and r'', and we take the the adversary in r'' to be $(N' \cup \{j''\} - \{j'\}, F'')$, where F'' interchanges the roles of j' and j''. We leave details to the reader. □

Theorem 6.5. *If $t \leq n-2$ then P^{min} implements $\mathbf{P}^0$ in the EBA context $\gamma_{\min,n,t}$.*

Proof. Let $\mathcal{I}$ be the system $\mathcal{I}_{\gamma_{\min,n,t},\mathbf{P}^0}$. We show that for all runs r and times m, we have $P_i^{min}(r_i(m)) = (\mathbf{P}_i^0)^{\mathcal{I}}(r_i(m))$.

- If $P_i^{min}(r_i(m)) = \mathsf{noop}$ because $decided_i \neq \perp$ in $r_i(m)$, we clearly also have $(\mathbf{P}_i^0)^{\mathcal{I}}(r_i(m)) = \mathsf{noop}$, because $\mathcal{I}, (r, m) \models K_i(decided_i \neq \perp)$.

- If $P_i^{min}(r_i(m)) = \mathtt{decide}_i(0)$, we must have $decided_i = \bot$ in $r_i(m)$, and either $init_i = 0$ or $jd_i = 0$. If $init_i = 0$ in $r_i(m)$, then $\mathcal{I}, (r, m) \models K_i(init_i = 0)$, so $(\mathbf{P}_i^0)^{\mathcal{I}}(r_i(m)) = \mathtt{decide}_i(0)$. If $jd_i = 0$ and $init_i \neq 0$ in $r_i(m)$, then we must have $time_i > 0$ in $r_i(m)$ and $jd_i = \bot$ in $r_i(m')$ for $m' < m$, for otherwise, agent i would have decided 0 earlier and we would have $decided_i \neq \bot$ in $r_i(m)$. Moreover, i must have received the message 0 from some agent j. Thus, $\mathcal{I}, (r, m) \models K_i(\bigvee_{j \in Agt} jdecided_j = 0)$. It follows that $(\mathbf{P}_i^0)^{\mathcal{I}}(r_i(m)) = \mathtt{decide}_i(0)$.
- If $P_i^{min}(r_i(m)) = \mathtt{decide}_i(1)$, we must have $decided_i = \bot$, $init_i = 1, jd_i \neq 0$, and $time_i = t + 1$ in $r_i(m)$. As shown in the argument for Termination in the proof of Proposition 6.1, we have $\mathcal{I}, (r, m) \models K_i(\bigwedge_{j \in Agt} \neg(jdecided_j = 0))$. Thus, $(\mathbf{P}_i^0)^{\mathcal{I}}(r_i(m)) = \mathtt{decide}_i(1)$.
- Finally, if $P_i^{min}(r_i(m)) = \mathtt{noop}$ by the final line of P_i^{min}, we must have $time_i < t + 1$, $init_i = 1$, $decided_i = \bot$, and $jd_i = \bot$ in $r_i(m)$. (Note that $t' = m$.) Consider a run r' where all agents have initial preference 1 and are nonfaulty. It is easy to see that i receives the same messages up to time m in r and r', so $r_i(m) = r_i'(m)$. Hence, we must have $\mathcal{I}, (r, m) \models \neg K_i(init_i = 0 \vee \bigvee_{j \in Agt} decided_j = 0)$. Thus, $(\mathbf{P}_i^0)^{\mathcal{I}}(r_i(m)) \neq \mathtt{decide}_i(0)$. It is also not hard to construct a run r'' that i considers possible where there are exactly t' faulty agents such that some nonfaulty agent j in r'' gets a 0-chain of length t' in round t' of r''. Thus, $\mathcal{I}, (r, m) \models \neg K_i(\bigwedge_{j \in Agt} \neg(deciding_j = 0))$, so $(\mathbf{P}_i^0)^{\mathcal{I}}(r_i(m)) \neq \mathtt{decide}_i(1)$. Therefore, $(\mathbf{P}_i^0)^{\mathcal{I}}(r_i(m)) = \bot$.

It follows that P^{min} implements $\mathbf{P}^0$ in $\gamma_{\min,n,t}$. □

Theorem 6.6. *If $t \leq n - 2$, then P^{basic} implements $\mathbf{P}^0$ in the EBA context $\gamma_{basic,n,t}$.*

Proof. We proceed just as in the previous argument. If $P_i^{basic}(r_i(m)) = \bot$ or $P_i^{basic}(r_i(m)) = \mathtt{decide}_i(0)$, then the argument is identical to that for P^{min}. If $P_i^{basic}(r_i(m)) = \mathtt{decide}_i(1)$, then we proceed by induction to show that when we have $decided_i = \bot$, $init_i = 1$, and either $jd_i = 1$ or $\#1_i > n - m$ in $r_i(m)$, then we also have $\mathcal{I}, (r, m) \models K_i(\bigwedge_{j \in Agt} \neg(deciding_j = 0))$.

So suppose that $decided_i = \bot$ and $init_i = 1$ are in $r_i(m)$. If $\#1_i > n - m$ in $r_i(m)$, then it is easy to see that there cannot be a 0-chain of length m in r (since the only agents that can be involved in this 0-chain are ones that did not send an $(init, 1)$ message). Thus, $\mathcal{I}, (r, m) \models K_i(\bigwedge_{j \in Agt} \neg(deciding_j = 0))$. On the other hand, if $jd_i = 1$ is in $r_i(m)$, then we must have that $jd_i = \bot$ in $r_i(m')$ for $m' < m$, for otherwise, agent i would have decided 1 earlier and we would have $decided_i \neq \bot$ in $r_i(m)$. Moreover, i must have received the message 1 from some agent j in round m of r. Thus, j decides 1 in round m of r, so we must have $decided_j = \bot$, $init_j = 1$, and either $jd_j = 1$ or $\#1_j > n - m$ in $r_j(m - 1)$. By the inductive hypothesis, $\mathcal{I}, (r, m - 1) \models K_j(\bigwedge_{j' \in Agt} \neg(deciding_{j'} = 0))$. If some agent j' decides 0 in round $m + 1$ of r, then there must be a 0-chain that ends 0 with j', so j' must get a message from an agent that decides 0 in round m, contradicting the fact that $\mathcal{I}, (r, m - 1) \models K_j(\bigwedge_{j' \in Agt} \neg(deciding_{j'} = 0))$. It thus follows from the information in $r_(m)$ that no agent decides 0 in round $m + 1$ of r, so $\mathcal{I}, (r, m) \models K_i(\bigwedge_{j \in Agt} \neg(deciding_j = 0))$. In either case, we have $(\mathbf{P}_i^0)^{\mathcal{I}}(r_i(m)) = \mathtt{decide}_i(1)$.

Finally, if $P_i^{basic}(r_i(m)) = \mathtt{noop}$ by the final line of P_i^{basic}, then arguments similar in spirit to those used above show that agent i considers it possible that all agents started with an initial preference of 1, and hence does not know that there is a 0-chain, but also cannot rule out the possibility of a 0-chain. We must have $decided_i = jd_i = \bot$, $init_i = 1$, and $\#1_i \leq n - m$ in $r_i(m)$. Since $init_i = 1$ and $jd_i = \bot$ are in $r_i(m)$, at the point (r, m), i considers possible the run r' where every agent started with an initial preference of 1 and the message pattern is identical to r. We then have $\mathcal{I}, (r, m) \models \neg K_i(\bigwedge_{j \in Agt}(jdecided_j = 0))$, so $(\mathbf{P}_i^0)^{\mathcal{I}}(r_i(m)) \neq \mathtt{decide}_i(0)$. Similarly, since $\#1_i \leq n - m$, i considers it possible that a there is a 0-chain consisting of the agents that i didn't hear a 1 from. Hence, $\mathcal{I}, (r, m) \models \neg K_i(\bigwedge_{j \in Agt} \neg(deciding_j = 0))$, so $(\mathbf{P}^0_i)^{\mathcal{I}}(r_i(m)) \neq \texttt{decide}_i(1)$. It follows that $(\mathbf{P}^0_i)^{\mathcal{I}}(r_i(m)) = \bot$. Thus, P^{basic} implements $\mathbf{P}^0$ in $\gamma_{basic,n,t}$. □

A.2 Proofs for Section 7

A.2.1 Motivating the KBP.

Proposition 7.2. *If P is an optimal protocol in the context $\gamma_{fip,n,t}$ and $\mathcal{I}_{P,\gamma_{fip,n,t}}, (r, m) \models decided_i = \bot \wedge K_i(C_{\mathcal{N}}(t\text{-}faulty \wedge no\text{-}decided_{\mathcal{N}}(1) \wedge \exists 0))$, then all undecided agents in $\mathcal{N}(r)$ make a decision in round $m + 1$, and similarly if $\mathcal{I}_{P,\gamma}, (r, m) \models decided_i = \bot \wedge K_i(C_{\mathcal{N}}(t\text{-}faulty \wedge no\text{-}decided_{\mathcal{N}}(0) \wedge \exists 1))$.*

Proof. For ease of exposition, let $\gamma = \gamma_{fip,n,t}$. Suppose that $\mathcal{I}_{P,\gamma}, (r, m) \models decided_i = \bot \wedge K_i(C_{\mathcal{N}}(t\text{-}faulty \wedge no\text{-}decided_{\mathcal{N}}(1) \wedge \exists 0))$. Let P' be the protocol for context γ obtained by modifying P as follows. For each agent i and local state $s \in L_i$, we define $P'_i(s)$ to be $\texttt{decide}_i(0)$ if $\mathcal{I}_{P,\gamma}, (r, m) \models K_i(C_{\mathcal{N}}(t\text{-}faulty \wedge no\text{-}decided_{\mathcal{N}}(1) \wedge \exists 0))$ for points (r, m) with $r_i(m) = s$, and $P'_i(s) = P_i(s)$ otherwise. We claim that $P' \leq P$ and P' is an EBA-protocol in context γ.

Let r and r' be corresponding runs of $\mathcal{I}_{P,\gamma}$ and $\mathcal{I}_{P',\gamma}$, respectively. We show by induction on k that $r(k) = r'(k)$ if k is less than or equal to the earliest time m such that $\mathcal{I}_{P,\gamma}, (r, m) \models C_{\mathcal{N}}(t\text{-}faulty \wedge no\text{-}decided_{\mathcal{N}}(1) \wedge \exists 0)$. The base case is trivial, and the inductive case follows the fact that if $r(k) = r'(k)$ and $\mathcal{I}_{P,\gamma}, (r, k) \not\models C_{\mathcal{N}}(t\text{-}faulty \wedge no\text{-}decided_{\mathcal{N}}(1) \wedge \exists 0))$, then for all agents i, $\mathcal{I}_{P,\gamma}, (r, k) \not\models K_i C_{\mathcal{N}}(t\text{-}faulty \wedge no\text{-}decided_{\mathcal{N}}(1) \wedge \exists 0)$ (the fact that $\mathcal{I}_{P,\gamma}, (r, k) \not\models K_i C_{\mathcal{N}}(t\text{-}faulty \wedge no\text{-}decided_{\mathcal{N}}(1) \wedge \exists 0)$ for all agents i follows from Lemma A.3, proved below), so $P_i(r_i(k)) = P'_i(r'_i(k))$ for all agents i. Since the failure patterns are the same in these runs, it follows that $r(k + 1) = r'(k + 1)$. Moreover, since $\mathcal{I} \models (C_{\mathcal{N}}(\phi) \wedge i \in \mathcal{N}) \Rightarrow K_i(C_{\mathcal{N}}(\phi))$ (see [5]), it follows that once $\mathcal{I}_{P,\gamma}, (r, m) \models C_{\mathcal{N}}(t\text{-}faulty \wedge no\text{-}decided_{\mathcal{N}}(1) \wedge \exists 0)$, all undecided agents in $\mathcal{N}(r')$ decide 0 simultaneously using P'.

It is immediate from these facts that $P' \leq_\gamma P$. If a nonfaulty agent i decides in r' at a time m before the common knowledge condition has become true in r, then $r(m) = r'(m')$ and agent i makes the same decision at time m in r. Once the common knowledge condition becomes true, an undecided nonfaulty agent i decides using P', so does so at least as soon as it does using P. Recall that if i is faulty, then the definition of $P' \leq_\gamma P$ allows that agent i decides using P before it does so using P', so we do not need to consider this case.

Next, we show that P' is an EBA protocol in context γ. Unique Decision follows from the fact that the context γ records decisions in the local state, and $P_i(s) = \bot$ for states s that record that a decision has already been made. For Validity, consider runs r' of P' and the corresponding run r of P in context γ. If the common knowledge condition has not yet become true at a point where nonfaulty agent i makes its decision on value v using P', then $r(m) = r'(m)$, and i makes the same decision using P. It follows from Validity for P that some agent has initial value v. Alternately, if nonfaulty agent i decides 0 at (r', m) because it knows at (r, m) in $\mathcal{I}_{P,\gamma}$ that the common knowledge condition has become true, then in fact $\mathcal{I}_{P,\gamma}, (r, m) \models \exists 0$, and the same fact holds at (r', m).

For Agreement, consider a run r' of P' where agent $i \in \mathcal{N}(r')$ decides 0 in round $m_0 + 1$ and agent $j \in \mathcal{N}(r')$ decides 1 in round $m_1 + 1$. If the common knowledge condition has not become true in the corresponding run r of P by time $\max(m_0, m_1)$, then since r and r' are identical up to at least this time, we have a contradiction to Agreement for P. If the common knowledge condition becomes true at time $k \leq m_1$, then according to P', all nonfaulty agents undecided by time k decide 0 in round $k + 1$. That means that agent j, which is undecided at time k, decides 0 in round k of r' and decides on a different value in round $m_1 + 1$, contradicting the fact that, as we have shown, P' satisfies Unique Decision. On the other hand, the common knowledge condition cannot become true after time m_1, since it implies $no\text{-}decided_{\mathcal{N}}(1)$ and $j \in \mathcal{N}(r')$ has decided 1 in round $m_1 + 1$. Thus, P' satisfies Agreeement($\mathcal{N}$).

Finally, Termination for P' follows from $P' \leq_\gamma P$ and the fact that P satisfies Termination.

Since we have $P' \leq P$ and P' is an EBA protocol for context γ, it follows that $P \leq P'$, so in $\mathcal{I}_{P,\gamma}$, all undecided nonfaulty agents make a decision as soon as they know that the common knowledge condition holds. □

Note that the result does not tell us *what* decision the nonfaulty agents must make when the common knowledge condition holds. There may be situations where no nonfaulty agent has made a decision, and both common knowledge conditions hold. Either a 0 or a 1 decision would be acceptable in this case.

We now prove all the required properties of $\mathbf{P}^1$. We start by examining what each of the conditions is $\mathbf{P}^1$ tells us.

A.2.2 The common knowledge conditions. As a first step to understanding $\mathbf{P}^1$, we characterize the common knowledge conditions (i.e., the second and third lines). The next proposition gives necessary and sufficient conditions for these conditions to hold. These conditions show that the problem of computing when the common knowledge conditions hold is tractable.

Define $dist_{\mathcal{N}}(t\text{-}faulty)$ to be an abbreviation for

$$\exists A \subseteq Agt(|A| = t \wedge \forall i \in A\, \exists j \in \mathcal{N}\, (K_j(i \notin \mathcal{N}))).$$

Roughly speaking, $dist_{\mathcal{N}}(t\text{-}faulty)$ holds if, between them, the nonfaulty agents know about t faulty agents.

*Definition A.1 (*hears-from*).* We define a *one-step hears-from* relation in a run r on pairs (j, m) consisting of agents j and times m by saying that (j', m') *one-step hears from* (j, m) in r if agent j sends j' a non-$\bot$ message in round $m+1$ of r that j'' receives and $m + 1 \leq m'$. The *hears-from* relation is the transitive closure of the *one-step hears-from* relation. We write $(j, m) \rightarrow_r (j', m')$ if (j', m') *hears-from* (j, m)

Those familiar with the *Lamport causality* relation [11] will recognize that the *hears-from* relation is similar in spirit.

Proposition A.2. *For all implementations P of the knowledge-based program $\mathbf{P}^1$ with respect to $\gamma_{fip,n,t}$,*

(a) $\mathcal{I}_{\gamma_{fip,n,t},P} \models time > 0 \Rightarrow (\ominus dist_{\mathcal{N}}(t\text{-}faulty) \Leftrightarrow C_{\mathcal{N}}(t\text{-}faulty))$.
(b) $\mathcal{I}_{\gamma_{fip,n,t},P} \models time > 0 \Rightarrow (no\text{-}decided_{\mathcal{N}}(v) \Rightarrow \wedge_{j \in \mathcal{N}} \ominus (K_j \bigcirc no\text{-}decided_j(v)))$ *and* $\mathcal{I}_{\gamma_{fip,n,t},P} \models (C_{\mathcal{N}}(t\text{-}faulty) \wedge (no\text{-}decided_{\mathcal{N}}(v))) \Leftrightarrow C_{\mathcal{N}}(t\text{-}faulty \wedge no\text{-}decided_{\mathcal{N}}(v))$, *for* $v \in \{0, 1\}$.
(c) $\mathcal{I}_{\gamma_{fip,n,t},P} \models time > 0 \Rightarrow ((C_{\mathcal{N}}(t\text{-}faulty) \wedge \ominus(\vee_{j \in \mathcal{N}} K_j(\exists v))) \Leftrightarrow C_{\mathcal{N}}(t\text{-}faulty \wedge \exists v))$, *for* $v \in \{0, 1\}$.

Proof. Let P be an implementation of $\mathbf{P}^1$ and let $\mathcal{I} = \mathcal{I}_{\gamma_{fip,n,t},P}$.

For part (a), first suppose by way of contradiction that $\mathcal{I}, (r, m) \models \neg \ominus dist_{\mathcal{N}}(t\text{-}faulty) \wedge C_{\mathcal{N}}(t\text{-}faulty)$. Let $A' = \{i : \exists j \in \mathcal{N}(r)(\mathcal{I}, (r, m-1) \models K_j(i \notin \mathcal{N}))\}$. By assumption, $|A'| < t$. There must exist some set A with $|A| = t$ such that $\mathcal{I}, (r, m) \models C_{\mathcal{N}}(\bigwedge_{i \in A}(i \notin \mathcal{N}))$. Moreover, A' is a strict subset of A, since each nonfaulty agent in r will learn in round m (if they did not already know it) that each agent in A' is faulty (by getting a message from a nonfaulty agent from which it can infer this). Let r' be a run where $\mathcal{N}(r') = A'$, all agents have the same initial state in r and r', and for all agents in A', the same messages are delivered in r and r'. It is easy to check that for all agents in $\mathcal{N}(r)$, we must have $(r, m-1) \sim_j (r', m-1)$. (Formally, we show by induction on k that if $j \in \mathcal{N}(r)$ and $k < m' \leq m-1$, then $(j', m'-k) \rightarrow_r (j, m')$ iff $(j', m'-k) \rightarrow_{r'} (j, m')$. That is, all agents in $\mathcal{N}(r)$ consider possible a run, namely r', where the only nonfaulty agents are those in A'. Now even though i may learn about other other nonfaulty agents in round m of r, for all $j \in \mathcal{N}(r)$

such that $j \neq i$, i must consider it possible that at (r, m), j considers (r', m) possible, because j hears from all agents other than those in A' in round m. Thus, *t-faulty* is not commnon knowledge among the nonfaulty agents at (r, m). (Note that here we are using the fact that $n - t \geq 2$, so that there is an agent $j \in \mathcal{N}(r)$ such that $j \neq i$.)

Conversely, suppose that $\mathcal{I}, (r, m-1) \models \textit{dist}_{\mathcal{N}}(\textit{t-faulty})$. Thus, there exists some set A with $|A| = t$ such that $\mathcal{I}, (r, m-1) \models \textit{dist}_{\mathcal{N}}(\textit{t-faulty}_A)$, where $\textit{dist}(\textit{t-faulty}_A)$ is the formula $\forall i \in A \exists j \in \mathcal{N}(K_j(i \notin \mathcal{N})))$. Moreover, for all runs r', if $\mathcal{I}, (r', m-1) \models \textit{dist}_{\mathcal{N}}(\textit{t-faulty}_A)$, then $\mathcal{N}(r) = Agt - A$.

It is well known (see [5]) that, for all formulas ϕ and ψ, if $\mathcal{I} \models \phi \Rightarrow E_{\mathcal{N}}(\psi \wedge \phi)$, then $\mathcal{I} \models \phi \Rightarrow C_{\mathcal{N}}\psi$. Thus, it suffices to show $\mathcal{I} \models \ominus\textit{dist}_{\mathcal{N}}(\textit{t-faulty}_A) \Rightarrow E_{\mathcal{N}}(\textit{t-faulty} \wedge \ominus\textit{dist}_{\mathcal{N}}(\textit{t-faulty}_A))$.

For all points (r', m'), if $\mathcal{I}, (r', m') \models \ominus\textit{dist}_{\mathcal{N}}(\textit{t-faulty}_A)$, then at (r', m), all the nonfaulty agents know that the agents in A are faulty and that these are the only faulty agents. Since $|A| = t$, we have that $\mathcal{I}, (r', m') \models E_{\mathcal{N}}(\textit{t-faulty})$. Moreover, since all the nonfaulty agents hear from all the other nonfaulty agents in round m of r', they all know $\ominus\textit{dist}_{\mathcal{N}}(\textit{t-faulty}_A)$. This completes the proof of part (a).

For part (b), the first part is immediate: if a nonfaulty agent i does not decide on a value v by round m, then at time $m - 1$, i knows this will be the case. The if direction of the second part is immediate from the fact that, since $\mathcal{N} \neq \emptyset$, $\mathcal{I} \models C_{\mathcal{N}}\phi \Rightarrow \phi$ for all formulas ϕ [5]. For the only-if direction, suppose that $\mathcal{I}, (r, m) \models C_{\mathcal{N}}(\textit{t-faulty}) \wedge \textit{no-decided}_{\mathcal{N}}(v)$. For each agent $i \in \mathcal{N}(r)$, as we obseved, i knows at time $m - 1$ in r that it will not decide v in round m. Since we are using a full-information protocol, all the nonfaulty agents will know at time m in r that $\textit{no-decided}_i(v)$ holds. Thus, $\mathcal{I}, (r, m) \models E_{\mathcal{N}}(\textit{no-decided}_{\mathcal{N}}(v))$. It is a standard property of $C_{\mathcal{N}}$ that $\mathcal{I} \models C_{\mathcal{N}}\phi \Rightarrow E_{\mathcal{N}}C_{\mathcal{N}}\phi$ [5]. Thus, $\mathcal{I}, (r, m) \models E_{\mathcal{N}}(\textit{no-decided}_{\mathcal{N}}(v) \wedge C_{\mathcal{N}}\textit{t-faulty})$. We have just shown that $\mathcal{I} \models C_{\mathcal{N}}(\textit{t-faulty}) \wedge \textit{no-decided}_{\mathcal{N}}(v)) \Rightarrow E_{\mathcal{N}}C_{\mathcal{N}}(\textit{t-faulty}) \wedge \textit{no-decided}_{\mathcal{N}}(v))$. It follows that $\mathcal{I} \models (C_{\mathcal{N}}(\textit{t-faulty}) \wedge (\textit{no-decided}_{\mathcal{N}}(v))) \Rightarrow C_{\mathcal{N}}(\textit{no-decided}_{\mathcal{N}}(v))$, as desired.

For part (c), the proof of the only-if direction is similar in spirit to that of part (b); we leave details to the reader. For the if direction, suppose that $\mathcal{I}, (r, m) \models C_{\mathcal{N}}(\textit{t-faulty}) \wedge \exists v$ and, by way of contradiction, $\mathcal{I}, (r, m) \not\models \vee_{j \in \mathcal{N}}(\ominus K_j(\exists v))$. Let r' be a run where all the agents have initial state $1 - v$, $\mathcal{N}(r') = \mathcal{N}(r)$, and the failure pattern is the same in r and r', except that in round m of r', all agents hear only from the nonfaulty agents. We claim that, for all $i \in \mathcal{N}(r)$, we have $(r, m-1) \sim_i (r', m-1)$. For clearly, if $j \in \mathcal{N}(r)$ and $k < m' \leq m - 1$, then $(j', m' - k) \rightarrow_r (j, m')$ iff $(j', m' - k) \rightarrow_{r'} (j, m')$. Since (j, m') does not hear from any agent with initial value v in r, it follows that this must also be the case in r'. Now a straightforward induction on m'' shows that if $j \in \mathcal{N}(r)$, $m'' \leq m' < m - 1$, and $(j', m'') \rightarrow_r (j, m')$, then $(r, m'') \sim_{j'} (r', m'')$. As in part (a), even though i may learn $\exists v$ in round m of r, for all $j \in \mathcal{N}(r)$ such that $j \neq i$, i must consider it possible that at (r, m), j considers (r', m) possible, so does not learn $\exists v$. Thus, $\exists v$ is not commnon knowledge among the nonfaulty agents at (r, m).

□

Lemma A.3. *For all implementations P of the knowledge-based program $\mathbf{P}^1$ with respect to $\gamma_{\mathit{fip},n,t}$, if $\mathcal{I}_{\gamma_{\mathit{fip},n,t},P}, (r, m) \models C_{\mathcal{N}}(\textit{t-faulty} \wedge \textit{no-decided}_{\mathcal{N}}(1 - v) \wedge \exists v)$, then for all $i \in Agt$, $\mathcal{I}_{\gamma_{\mathit{fip},n,t},P}, (r, m) \models K_i(C_{\mathcal{N}}(\textit{t-faulty} \wedge \textit{no-decided}_{\mathcal{N}}(1 - v) \wedge \exists v))$. Moreover, for all agents i, $\mathcal{I}_{\gamma_{\mathit{fip},n,t},P}, (r, m) \models C_{\mathcal{N}}(\textit{t-faulty}) \Rightarrow K_i(C_{\mathcal{N}}(\textit{t-faulty})$.*

Proof. Let $\mathcal{I} = \mathcal{I}_{\gamma_{\mathit{fip},n,t},P}$ and suppose that $\mathcal{I}, (r, m) \models C_{\mathcal{N}}(\textit{t-faulty} \wedge \textit{no-decided}_{\mathcal{N}}(1 - v) \wedge \exists v)$ for some $v \in \{0, 1\}$ and $(r, m) \sim_i (r', m)$. We clearly must have $m > 0$, since we cannot have common knowledge of the faulty agents at time 0. By Proposition A.2, it follows that $\mathcal{I}, (r, m-1) \models \textit{dist}_{\mathcal{N}}(\textit{t-faulty}) \wedge (\wedge_{j \in \mathcal{N}}(K_j \bigcirc \textit{no-decided}_j(v)) \wedge (\vee_{j \in \mathcal{N}} K_j(\exists v))$. Since the nonfaulty agents send messages to all agents in round m, all agents (not just the nonfaulty agents) will know at time m

that this was true at time $m-1$. It now follows from Proposition A.2 that, for all agents i, we have $\mathcal{I}_{\gamma_{\mathit{fip},n,t},P}, (r,m) \models K_i(C_{\mathcal{N}}(\mathit{t\text{-}faulty} \wedge \mathit{no\text{-}decided}_{\mathcal{N}}(1-v) \wedge \exists v))$. □

We next show that once the nonfaulty agents know about t faulty agents, every agent decides by the end of the following round.

Lemma A.4. *For all implementations P of the knowledge-based program $\mathbf{P}^1$ with respect to $\gamma_{\mathit{fip},n,t}$, $\mathcal{I}_{\gamma_{\mathit{fip},n,t},P} \models C_{\mathcal{N}}(\mathit{t\text{-}faulty}) \Rightarrow \bigwedge_{i\in Agt} \bigcirc\neg(\mathit{decided}_i = \bot)$.*

Proof. Let $\mathcal{I} = \mathcal{I}_{\gamma_{\mathit{fip},n,t},P}$ and suppose that $i \in Agt$, $j \in \mathcal{N}$, and $\mathcal{I},(r,m) \models C_{\mathcal{N}}(\mathit{t\text{-}faulty})$. If j decides 0 in round $m' \leq m$, then i must decide by round $m'+1$, because it will hear from j in round m' that it is deciding, so that $\mathcal{I},(r,m') \models K_i(j\mathit{decided}_j = 0)$. (Note that i may decide earlier or due to one of the common knowledge conditions, but it will surely decide by round $m'+1$.) If j decides 1 at some round $m' \leq m$, then either

- $\mathcal{I},(r,m'-1) \models K_j(C_{\mathcal{N}}(\mathit{t\text{-}faulty} \wedge \mathit{no\text{-}decided}_{\mathcal{N}}(0) \wedge \exists 1)$, so $\mathcal{I},(r,m'-1) \models K_i(C_{\mathcal{N}}(\mathit{t\text{-}faulty} \wedge \mathit{no\text{-}decided}_{\mathcal{N}}(0) \wedge \exists 1)$ by Lemma A.3, and i must also decide by round m';
- $\mathcal{I},(r,m'-1) \models K_j(\bigwedge_{k\in Agt}\neg(\mathit{deciding}_k = 0))$, so, since P is a full-information protocol, $\mathcal{I},(r,m') \models \ominus K_j(\bigwedge_{k\in Agt}\neg(\mathit{deciding}_k = 0))$. Since P is a FIP, i hears from j in round m' that no agent decides 0 in round m'. Thus, i knows that no agent will decide 0 in round $m'+1$ due to line 4 of P. If some agent decides 0 due to the common knowledge condition in round $m'+1$, then, by the argument above, i also decides 0 in round $m'+1$. If not, then line 5 of P applies, and i decides 1 in round $m'+1$.

We have just shown that that i decides by round $m+1$ if some nonfaulty agent j decides by round m. If no nonfaulty agent decides by round m, then $\mathcal{I},(r,m) \models \mathit{no\text{-}decided}_{\mathcal{N}}(0) \wedge \mathit{no\text{-}decided}_{\mathcal{N}}(1)$. We must have $m > 0$ since $C_{\mathcal{N}}(\mathit{t\text{-}faulty})$ cannot hold at time 0, and clearly $\mathcal{I},(r,m-1) \models (\vee_{k\in\mathcal{N}}\exists 0) \vee (\vee_{k\in\mathcal{N}}\exists 1)$. It now follows easily from Proposition A.2 that $\mathcal{I},(r,m) \models C_{\mathcal{N}}(\mathit{t\text{-}faulty} \wedge \mathit{no\text{-}decided}_{\mathcal{N}}(1) \wedge \exists 0)$ or $\mathcal{I},(r,m) \models C_{\mathcal{N}}(\mathit{t\text{-}faulty} \wedge \mathit{no\text{-}decided}_{\mathcal{N}}(0) \wedge \exists 1)$; that is, one of the common knowledge conditions must hold. Hence, by Lemma A.3, agent i must decide by round $m+1$. □

A.2.3 Characterizing the condition for deciding 0.

Lemma A.5. *For all implementations P of the knowledge-based program $\mathbf{P}^1$ with respect to $\gamma_{\mathit{fip},n,t}$, if $\mathcal{I}_{\gamma_{\mathit{fip},n,t},P}, (r,m) \models \neg K_i(C_{\mathcal{N}}(\mathit{t\text{-}faulty} \wedge \mathit{no\text{-}decided}_{\mathcal{N}}(1) \wedge \exists 0)) \wedge \neg K_i(C_{\mathcal{N}}(\mathit{t\text{-}faulty} \wedge \mathit{no\text{-}decided}_{\mathcal{N}}(0) \wedge \exists 1))$, then agent i receives a 0-chain in round m if and only if $\mathcal{I}_{\gamma_{\mathit{fip},n,t},P}, (r,m) \models \mathit{init}_i = 0 \vee \bigvee_{j\in Agt} K_i(j\mathit{decided}_j = 0)$ and i has not decided before round $m+1$.*

Proof. We proceed by induction on m. Let $\mathcal{I} = \mathcal{I}_{\gamma_{\mathit{fip},n,t},P}$ and suppose that $\mathcal{I},(r,m) \models \neg K_i(C_{\mathcal{N}}(\mathit{t\text{-}faulty} \wedge \mathit{no\text{-}decided}_{\mathcal{N}}(1) \wedge \exists 0)) \wedge \neg K_i(C_{\mathcal{N}}(\mathit{t\text{-}faulty} \wedge \mathit{no\text{-}decided}_{\mathcal{N}}(0) \wedge \exists 1))$. By Lemma A.3, we then have $\mathcal{I},(r,m) \models \neg C_{\mathcal{N}}(\mathit{t\text{-}faulty} \wedge \mathit{no\text{-}decided}_{\mathcal{N}}(1) \wedge \exists 0) \wedge \neg C_{\mathcal{N}}(\mathit{t\text{-}faulty} \wedge \mathit{no\text{-}decided}_{\mathcal{N}}(0) \wedge \exists 1)$.

If $m = 0$, then the only-if direction follows from the fact that agent i receives a 0-chain in round 0 iff $\mathcal{I},(r,0) \models \mathit{init}_i = 0$. Obviously, i has not decided at time 0. For the converse, we first observe that $\mathcal{I},(r,0) \models \neg K_i(j\mathit{decided}_j = 0)$ for all $j \in Agt$ as at time 0, $\mathcal{I},(r,0) \models \neg(j\mathit{decided}_j = 0)$ for all $j \in Agt$. Thus, $\mathcal{I}_{\gamma_{\mathit{fip},n,t},P}, (r,0) \models \mathit{init}_i = 0 \vee \bigvee_{j\in Agt} K_i(j\mathit{decided}_j = 0)$ iff $\mathcal{I}_{\gamma_{\mathit{fip},n,t},P}, (r,0) \models \mathit{init}_i = 0$. So agent i decides 0 in round 1 and receives a 0-chain in round 0.

If $m > 0$, then the only-if direction follows immediately from the definition of a 0-chain, as the last agent i decides for the first time after hearing from an agent that just decided 0. For the converse, suppose that $\mathcal{I},(r,m) \models \mathit{init}_i = 0 \vee \bigvee_{j\in Agt} K_i(j\mathit{decided}_j = 0)$ holds for the first time at time m. Since $m > 0$ is the first time this formula holds and i did not decide before round $m+1$, we must have $\mathcal{I},(r,m) \models \mathit{init}_i \neq 0$. Thus, $\mathcal{I},(r,m) \models K_i(j\mathit{decided}_j = 0)$ for some $j \in Agt$, so agent j decides 0 in round m. We must have $\mathcal{I},(r,m-1) \models \neg C_{\mathcal{N}}(\mathit{t\text{-}faulty})$ since otherwise, by Lemma A.4,

i would have decided by round m. Thus, j could not have decided 0 due to the common knowledge condition. It follows from $\mathbf{P}^1$ that $\mathcal{I}, (r, m-1) \models init_j = 0 \vee \bigvee_{j' \in Agt} K_j(jdecided_{j'} = 0)$. By the inductive hypothesis, j receieves a 0-chain in round $m-1$. Since $\mathcal{I}, (r, m) \models K_i(jdecided_j = 0)$, agent i receives the message sent by j in round m. Because the common knowledge conditions are not satisfied and i has not already decided, agent i also decides 0 in round $m+1$. Thus. i receives a 0-chain in round $m+1$. □

A.2.4 Characterizing the condition for deciding 1. To characterize when an agent is unable to decide 1 (excluding the decisions made using one of the common knowledge conditions), we need some additional definitions.

Definition A.6. Let $len_i(r, m)$ be the length of the longest 0-chain that i knows about at time m in run r (where $len_i(r, m) = 0$ if i does not know about any 0-chains), let $last_{ij}(r, m)$ be the last time m' for which $(j, m') \longrightarrow_r (i, m)$ (where $last_{ij}(r, m) = -1$ if (i, m) does not hear from (j, m') at all) and let $latest0_i(r, m)$ to be the last time m' such that, for some agent j, we have $(j, m') \longrightarrow_r (i, m)$ and $\mathcal{I}, (r, m') \models deciding_j = 0$ (where $latest0_i(r, m) = -1$ if there is no such time).

Intuitively, if the common knowledge conditions do not hold by time m, then an agent i is unable to decide 1 if and only if there are enough agents that could potentially extend the longest 0-chain that i knows about, so that it has length at least m. The following lemma formalizes this intuition.

Proposition A.7. *For all implementations P of the knowledge-based program $\mathbf{P}^1$ with respect to $\gamma_{fip,n,t}$, if $\mathcal{I}_{\gamma_{fip,n,t},P}, (r, m) \models \neg K_i(C_{\mathcal{N}}(t\text{-}faulty \wedge no\text{-}decided_{\mathcal{N}}(0) \wedge \exists 1))$, $\mathcal{I}_{\gamma_{fip,n,t},P}, (r, m) \models \neg K_i(C_{\mathcal{N}}($ $t\text{-}faulty \wedge no\text{-}decided_{\mathcal{N}}(1) \wedge \exists 0))$, and $\mathcal{I}, \gamma_{fip,n,t}, P, (r, m) \models decided_i = \bot$, then the following holds:*

- *$\mathcal{I}_{\gamma_{fip,n,t},P}, (r, m) \models \neg K_i \neg(\exists j \in Agt(deciding_j = 0))$ if and only if for all m'' with $len_i(r, m) < m'' \leq m$, there exist at least $m'' - len_i(r, m)$ agents j such that $last_{ij}(r, m) < m''$ and $\mathcal{I}, (r, last_{ij}(r, m) + 1) \models decided_j = \bot$.*

Proof. Let $\mathcal{I} = \mathcal{I}_{\gamma_{fip,n,t},P}$ and suppose that $\mathcal{I}, (r, m) \models \neg K_i(C_{\mathcal{N}}(t\text{-}faulty \wedge no\text{-}decided_{\mathcal{N}}(1) \wedge \exists 0)) \wedge \neg K_i(C_{\mathcal{N}}(t\text{-}faulty \wedge no\text{-}decided_{\mathcal{N}}(0) \wedge \exists 1))$. By Lemma A.3, we have that $\mathcal{I}, (r, m) \models \neg C_{\mathcal{N}}(t\text{-}faulty \wedge no\text{-}decided_{\mathcal{N}}(1) \wedge \exists 0) \wedge \neg C_{\mathcal{N}}(t\text{-}faulty \wedge no\text{-}decided_{\mathcal{N}}(0) \wedge \exists 1)$. Note that the common knowledge conditions also do not hold for any earlier time, as otherwise i would have decided earlier by Lemma A.4.

For the only-if direction, suppose that $\mathcal{I}, (r, m) \models \neg K_i \neg(deciding_j = 0)$ for some $j \in Agt$. There must exist a run r' such that $(r, m) \sim_i (r', m)$ and $\mathcal{I}, (r', m) \models deciding_j = 0$.

Since $(r, m) \sim_i (r', m)$ and i the common knowledge conditions do not hold for i in (r, m), they do not hold in (r', m) either. By Lemma A.5, j must receive a 0-chain in round m of r'. Let $i_0, \ldots, i_m$ be this 0-chain. For all m^* such that $len_i(r, m) < m^* \leq m$, consider the agent i_{m^*} in the 0-chain. We claim that $(i_{m^*}, m') \not\rightarrow_r (i, m)$ for all $m' \geq m^*$. For suppose that $(i_{m^*}, m') \longrightarrow_r (i, m)$. Then, since $(r, m) \sim_i (r', m)$, we also have $(i_{m^*}, m') \longrightarrow_{r'} (i, m)$. Since i_{m^*} receives a 0-chain at time m^* in r', this must also be the case in r and i learns about it, contradicting the assumption that $len_i(r, m) < m^*$. We must have $last_{ii_{m^*}}(r, m) < m^*$ (otherwise $len_i(r, m)$ would be at least m^*). It follows that $\mathcal{I}, (r, last_{ii_{m^*}}(r, m) + 1) \models decided_{i_m^*} = \bot$, for otherwise i_{m^*} would not be on the 0-chain. Thus, for all m'' with $len_i(r, m) < m'' \leq m$, there exist at least $m'' - len_i(r, m)$ agents j such that $last_{ij}(r, m) < m''$, namely, the agents i_{m^*} with $len_i(r, m) < m^* \leq m^*$.

Conversely, suppose that for all m'' where $len_i(r, m) < m'' \leq m$, there exist at least $m'' - len_i(r, m)$ agents j such that $last_{ij} < m''$ and $\mathcal{I}, (r, last_{ij}(r, m) + 1) \models decided_j = \bot$. Let $i_0, i_1, \ldots, i_{len_i(r,m)}$ be the longest 0-chain that i knows about at time m in run r.

By assumption, there must exist agents $i_{len_i(r,m)+1}, \ldots, i_m$ such that for all m'' with $len_i(r, m) < m'' \leq m$, we have $last_{ii_{m''}} < m''$ and $\mathcal{I}, (r, last_{ii_{m''}}(r, m) + 1) \models decided_{i_{m''}} = \bot$. Consider a run r''

such that the initial state of all agents is the same in r and r'', $\mathcal{N}(r) = \mathcal{N}(r'')$, and the failure pattern in r and r' is the same except for the messages received by and from the agents $i_{\mathit{len}_i(r,m)+1}, \dots, i_m$. Each agent $i_{m''}$ with $\mathit{len}_i(r,m) < m'' \leq m$ receives messages from the same agents in r and r'' up to and including round $\mathit{last}_{i,i_{m''}}$ and the same agents receive messages from $i_{m''}$ up to round $\mathit{last}_{i,i_{m''}}$; after round $\mathit{last}_{i,i_{m''}}$, agent $i_{m''}$ receives only one message, from agent $i{+}_{m''-1}$ at round m'', and if $m'' < m$, only one message is received from $i_{m''}$ after round $\mathit{last}_{i,i_{m''}}$: a message is received by $i_{m''+1}$ from $i_{m''}$ at round $m''+1$. (If $m'' = m$, no messages are received from $i_{m''}$ after round $\mathit{last}_{i,i_{m''}}$.) It is easy to check that $(r,m) \sim_i (r'',m)$ and $i_0, \dots, i_m$ is a 0-chain in r''. □

Corollary A.8. *For all implementations P of the knowledge-based program $\mathbf{P}^1$ with respect to $\gamma_{\mathit{fip},n,t}$, if $\mathcal{I}_{\gamma_{\mathit{fip},n,t},P}, (r,m) \models \neg C_{\mathcal{N}}(t\text{-}\mathit{faulty})$, $\mathcal{I}_{\gamma_{\mathit{fip},n,t},P}, (r,m) \models \mathit{decided}_i = \bot$, $\mathit{len}_i(r,m) \leq m-2$, and agent i hears from all but one agent in round m of r, then $\mathcal{I}_{\gamma_{\mathit{fip},n,t},P}, (r,m) \models \mathit{deciding}_i = 1$.*

Proof. Suppose that $\mathcal{I}_{\gamma_{\mathit{fip},n,t},P}, (r,m) \models \neg C_{\mathcal{N}}(t\text{-}\mathit{faulty})$, $\mathit{len}_i(r,m) \leq m-2$, and agent i hears from all but one agent in round m of r. For all agents j that i heard from in round m, $\mathit{last}_{ij}(r,m) \geq m$. So there is only one agent that could satisfy $\mathit{last}_{ij}(r,m) < m$. But since $m - \mathit{len}_i(r,m) \geq 2$, Proposition A.7 implies that $\mathcal{I}_{\gamma_{\mathit{fip},n,t},P}, (r,m) \models K_i \neg(\exists j \in \mathit{Agt}(\mathit{deciding}_j = 0))$. Therefore agent i decides 1 in round $m+1$. □

A.2.5 $\mathbf{P}^1$ satisfies the EBA conditions.

Proposition 7.3. *All implementations of $\mathbf{P}^1$ with respect to $\gamma_{\mathit{fip},n,t}$ are EBA decision protocols for $\gamma_{\mathit{fip},n,t}$.*

Proof. Fix an implementation P of $\mathbf{P}^1$ in $\gamma_{\mathit{fip},n,t}$.

Unique Decision follows from essentially the same argument as in the $\mathbf{P}^0$ case. The only difference is since P is a FIP, the decisions can be inferred from the local state without explicitly storing them.

To see that Agreement holds, suppose by way of contradiction that r is a run where there exist nonfaulty agents i and j and a time m such that $\mathcal{I}, (r,m) \models \mathit{decided}_i = 0 \land \mathit{decided}_j = 1$. Suppose that j decides 1 in round $m_j + 1$ and i decides 0 in round $m_i + 1$, so that the decision conditions first hold at times m_i and m_j, respectively. We first observe that if either i or j decides using one of the common knowledge conditions, then we get a contradiction. If $\mathcal{I}, (r,m_i) \models K_i(C_{\mathcal{N}}(t\text{-}\mathit{faulty} \land \mathit{no\text{-}decided}_{\mathcal{N}}(1) \land \exists 0))$ then j couldn't have decided 1 at or before round m_i, since $j \in \mathcal{N}$ and $\mathcal{I}, (r,m_i) \models \mathit{no\text{-}decided}_{\mathcal{N}}(1)$. Agent j would then decide 0 at round $m_i + 1$, since $\mathcal{I}, (r,m_i) \models K_j(C_{\mathcal{N}}(t\text{-}\mathit{faulty} \land \mathit{no\text{-}decided}(1) \land \exists 0))$, contradicting the assumption that j decides 1 in this run. If $\mathcal{I}, (r,m_i) \models \neg K_i(C_{\mathcal{N}}(t\text{-}\mathit{faulty} \land \mathit{no\text{-}decided}_{\mathcal{N}}(1) \land \exists 0))$ and $\mathcal{I}, (r,m_j) \models K_j(C_{\mathcal{N}}(t\text{-}\mathit{faulty} \land \mathit{no\text{-}decided}_{\mathcal{N}}(0) \land \exists 1))$, then i could not have decided 0 before m_j as $i \in \mathcal{N}$ and $\mathcal{I}, (r,m_j) \models \mathit{no\text{-}decided}_{\mathcal{N}}(0)$. Since $\mathcal{I}, (r,m_j) \models K_i(C_{\mathcal{N}}(t\text{-}\mathit{faulty} \land \mathit{no\text{-}decided}_{\mathcal{N}}(0) \land \exists 1))$, and the other common knowledge condition for deciding 0 does not hold for i at time m_j, by the argument above, i decides 1 in round $m_j + 1$, contradicting the assumption that i decides 0 in round $m_i + 1$.

It remains to consider the cases where neither i nor j decides using one of the common knowledge conditions.

If $m_j \leq m_i$, must have either $\mathcal{I}, (r,m_i) \models \mathit{init}_i = 0$ or $\mathcal{I}, (r,m_i) \models K_i(\mathit{jdecided}_k = 0)$ for some $k \in \mathit{Agt}$. Since, agent i does not decide using a common knowledge condition, we can apply Lemma A.5 to conclude that i receives a 0-chain at time m_i, which implies that there exists an agent i' such that $\mathcal{I}, (r,m_j) \models \mathit{deciding}_{i'} = 0$. Hence, $\mathcal{I}, (r,m_j) \models \neg K_j(\neg(\mathit{deciding}_{i'} = 0))$, so j cannot decide 1 at m_j.

If $m_j > m_i$, since i decides 0 in round $m_i + 1$ without using the common knowledge condition, we must have $\mathcal{I}, (r,m_i) \models \mathit{init}_i = 0 \bigvee_{j \in \mathit{Agt}} K_i(\mathit{jdecided}_j = 0)$. We can again apply Lemma A.5 to conclude that i receives a 0-chain at time m_i in r. As i is nonfaulty, j must hear from i in round $m_i + 1$, so $\mathcal{I}, (r, m_j) \models K_j(jdecided_i = 0)$. It follows that agent j should decide 0 in this run, contradicting the assumption that j decides 1.

For Validity, observe that if an agent i decides v using the common knowledge condition, it follows that some agent had an initial preference of v. If i decides 0 without using the common knowledge condition, by Lemma A.5, there must be a 0-chain, and hence an agent that had an initial preference of 0. Finally, if agent i decides 1 without using a common knowledge condition, then i did not decide 0 in the first round and therefore we must have $init_i = 1$.

For Termination, we must show that all nonfaulty agents decide by round $t + 2$. Suppose that a nonfaulty agent i does not decide by round $t + 1$ in run r, and that r' is a run such that $(r', t + 1) \sim_i (r, t + 1)$. Since we are using a full-information protocol, it easily follows that $(r, m) \sim_i (r', m)$ for all $m \leq t + 1$. Since i does not decide by round $t + 1$ of r, we do not have $\mathcal{I}, (r, m) \models K_i C_{\mathcal{N}}(\textit{t-faulty} \wedge \textit{no-decided}_{\mathcal{N}}(1 - v) \wedge \exists v)$ for $v \in \{0, 1\}$ and $m \leq t$. By Lemma A.3, it easily follows that $\mathcal{I}(r', m) \models \neg C_{\mathcal{N}}(\textit{t-faulty} \wedge \textit{no-decided}_{\mathcal{N}}(1 - v) \wedge \exists v)$ for $v \in \{0, 1\}$ and $m \leq t$. This implies that no agent j decides in r' using the common knowledge conditions at or before round $t + 1$.

By Lemma A.5, if any agent j decides 0 in round $t + 1$ in r', then that agent receives a 0-chain $i_0, \ldots, i_t$ at time t in r'. Since agents on a chain are distinct, and there are at most t faulty agents, this chain contains at least one nonfaulty agent i_k that decides 0. But then i must receive a message (and a 0-chain) from i_k in round $k + 1$, which means that i decides 0 by round $t + 2$, as claimed.

We have shown if an agent decides by round $t + 1$ using the commnon knowledge conditions or decides 0 by line 4 of $\mathbf{P}^1$ in a run r' that i considers possible, then i decides by round $t + 2$ in r. If this is not the case, then $\mathcal{I}, (r, t + 1) \models K_i(\bigwedge_{j \in Agt} \neg(deciding_j = 0))$, so i decides 1 in round $t + 2$. □

A.2.6 $\mathbf{P}^1$ *is optimal.* To show that $\mathbf{P}^1$ is optimal, we first show that it suffices to prove that weak safety implies optimality, and then show $\mathbf{P}^1$ is weakly safe.

Theorem 7.6. *If*$\mathbf{P}^1$ *is weakly safe with respect to* $\gamma_{\mathit{fip},n,t}$ *then all implementations of*$\mathbf{P}^1$ *are optimal with respect to* $\gamma_{\mathit{fip},n,t}$.

Proof. Suppose that $\mathbf{P}^1$ is weakly safe with respect to $\gamma_{\mathit{fip},n,t}$. To prove that all implementations of $\mathbf{P}^1$ are optimal with respect to $\gamma_{\mathit{fip},n,t}$, we use Theorem 7.5. It suffices to prove the only if direction, since Proposition 4.3 in [8] shows that the if direction holds for EBA protocols, and Proposition 7.3 shows that $\mathbf{P}^1$ is an EBA protocol.

Suppose that $\mathbf{P}^1$ is weakly safe with respect to $\gamma_{\mathit{fip},n,t}$. Let P be an implementation of $\mathbf{P}^1$ and let $\mathcal{I} = \mathcal{I}_{\gamma_{\mathit{fip},n,t},P}$, We give the argument for the $\mathcal{N} \wedge \mathcal{O}$ case. We first assume that $\mathcal{I}, (r, m) \models i \in \mathcal{N}$ for some point (r, m). In terms of $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$-reachability, we want to show that if for all points (r', m) such that $i \in \mathcal{N}(r')$, $\mathcal{I}, (r', m) \models \exists 0 \wedge \neg(\bigcirc(decided_i = 1))$, and for all points (r'', m') that are $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$-reachable from (r', m), we have $\mathcal{I}, (r'', m') \models \exists 0$, then $\mathcal{I}, (r, m) \models \bigcirc(decided_i = 0)$. Suppose by way of contradiction that for all (r', m) such that $i \in \mathcal{N}(r')$ the condition above holds but $\mathcal{I}, (r, m) \models \neg\bigcirc(decided_i = 0)$. Then, $\mathcal{I}, (r, m) \models i \in \mathcal{N} \wedge \bigcirc(decided_i = \bot)$ and by weak safety, there exist points $(r^{1'}, m)$ and $(r^{1''}, m)$ such that $r_i(m) = r^{1'}_i(m)$, $i \in \mathcal{N}(r^{1'})$, $(r^{1''}, m')$ is $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$-reachable from $(r^{1'}, m)$, and $\mathcal{I}, (r^{1''}, m') \models \neg\exists 0$. This is a contradiction. Since this holds for all (r, m), the only-if direction of the first optimality condition holds. The argument for the $\mathcal{N} \wedge \mathcal{Z}$ case is completely analogous.

Therefore all implementations P of $\mathbf{P}^1$ with respect to $\gamma_{\mathit{fip},n,t}$ are optimal. □

Let P be an implementation on $\mathbf{P}^1$ and let $\mathcal{I} = \mathcal{I}_{\gamma_{\mathit{fip},n,t},P}$.. We want to show that $\mathbf{P}^1$ is weakly safe. So suppose that $\mathcal{I}, (r, m) \models i \in \mathcal{N} \wedge \bigcirc(decided_i = \bot)$. We need to show that there exist points $(r^{0'}, m)$, $(r^{0''}, m)$, $(r^{1'}, m)$, and $(r^{1''}, m)$ satisfying the conditions of weak safety. Before we do this, we introduce an invariant.

Definition A.9 (Invariant condition for $v \in \{0,1\}$). $\mathcal{I}, (r,m) \models inv_v(i,j,k)$ if and only if there exist distinct agents i, j, k such that:

- $\mathcal{I}, (r,m) \models (decided_i = \bot) \land (deciding_j = v) \land (decided_k = \bot)$,
- $\mathcal{I}, (r,m) \models i \in \mathcal{N} \land j \in \mathcal{N} \land k \notin \mathcal{N}$, and
- k does not exhibit any faulty behaviour throughout r.

The motivation for the constraint on k in the definition is the following:

Lemma A.10. *If $k \notin \mathcal{N}(r)$ and k does not exhibit faulty behavior in r, then $\mathcal{I}, (r,m) \models \neg C_{\mathcal{N}}(t\text{-}faulty)$ for all m.*

Proof. If If $k \notin \mathcal{N}(r)$ but does not exhibit fauilty behavior, let r' be a run such that $\mathcal{N}(r') = \mathcal{N}(r) - \{k\}$, the failure pattern in r and r' is the same, and all agents have the same initial preferences in r and r'. Clearly, for all times m, $(r,m) \sim_i (r',m)$ and $\mathcal{I}, (r',m) \not\models t\text{-}faulty$, so $\mathcal{I}, (r,m) \models \neg C_{\mathcal{N}}(t\text{-}faulty)$. □

It follows immediately that $\mathcal{I}, (r,m) \models inv_1(i,j,k)$, then no agent decides using the common knowledge conditions in r.

We are now ready to construct the points $(r_i^{0''}, m)$ and $(r^{1''}, m)$ required for weak safety. The main part of the argument for this case is done by Lemmas A.15 and A.16. Lemma A.15 shows that once we are at a point where the invariant condition for 1 holds, we can $(\mathcal{N} \land \mathcal{O})$-$\boxdot$-reach a point where every agent has initial preference 1. Lemma A.16 shows that if a nonfaulty agent i is unable to decide, i must consider possible a run where the invariant condition for 1 holds. The desired result follows from these two lemmas.

The following technical lemma will play a key role in our proof of weak safety. For a point (r,m), we write $E(r,m)$ for the set of edges $(i, k-1) \rightarrow_r (j,k)$ with $1 \leq k \leq m$.

Lemma A.11. *If $i \in \mathcal{N}(r)$, $\mathcal{I}, (r,m) \models \neg C_{\mathcal{N}}(t\text{-}faulty)$, $latest0_i(r,m) = \ell < m$, and $\mathcal{I}, (r,m) \models decided_i = \bot$, then there exists a run r' such that*

- *$(r,m) \sim_i (r',m)$,*
- *$\mathcal{N}(r) = \mathcal{N}(r')$,*
- *$E(r',m) \subseteq E(r,m)$,*
- *all 0-chains in r' are known to agent i at time m; that is, if $i_0, \dots i_k$ is a 0-chain in r', then $(i_k, k) \rightarrow_{r'} (i,m)$,*
- *all agents that do not exhibit faulty behavior in r also do no not exhibit faulty behavior in r'.*

Proof. Note that since $\mathcal{I}, (r,m) \models decided_i = \bot$ and $latest0_i(r,m) < m$, we cannot have $\mathcal{I}, (r,m) \models deciding_i = 0$, so in fact agent i cannot hear from any agent that decides 0 in round m. This means that we must have $\ell = latest0_i(r,m) < m-1$. We construct r' by modifying r appropriately. For all $(j,m') \not\rightarrow_r (i,m)$, if $m' = 0$ and $\mathcal{I}, (r,0) \models init_j = 0$, then we modify $init_j$ to 1 in r', and if $m' > 0$, we restrict the messages arriving at (j,m') to be only those from the nonfaulty agents in $\mathcal{N}(r)$, and the agents that exhibit no faulty behavior in r. We have $\mathcal{N}(r') = \mathcal{N}(r)$.

Clearly, $E(r',m) \subseteq E(r,m)$. Since we do not modify the failure pattern for messages corresponding to pairs on the path $(j,m') \rightarrow_r (i,m)$, we have $(r,m) \sim_i (r',m)$. Moreover, the construction changes only the failure behaviour of agents who exhibit faulty behavior in r, so all agents that do not exhibit faulty behavior in r also do not exhibit faulty behavior in r'.

Note that because $\mathcal{I}, (r,m) \models i \in \mathcal{N} \land \neg C_{\mathcal{N}}(t\text{-}faulty)$ and $(r,m) \sim_i (r',m)$, we have $\mathcal{I}, (r',m) \models \neg C_{\mathcal{N}}(t\text{-}faulty)$, so in r', no agent decides using the common knowledge conditions before round $m+2$. In particular, any 0-decisions made before this round must be made using the fourth line of the knowledge-based program.

The construction guarantees that if $i_0, \dots, i_\ell$ is a 0-chain in r', then $(i_\ell, \ell) \rightarrow_{r'} (i, m)$. To see this, first note that none of $i_0, \dots, i_\ell$ can be nonfaulty, otherwise i would receive a 0-chain in r and r', and decide 0 before time m, contradicting the assumption that $\mathcal{I}, (r, m) \models \mathit{deciding}_i = \bot$. If $\ell > 0$, then the fact that i_ℓ received a message from $i_{\ell-1}$ in r' and $i_{\ell-1}$ is faulty means that $(i_\ell, \ell) \rightarrow_r (i, m)$. It easily follows that $(i_\ell, \ell) \rightarrow_{r'} (i, m)$. Thus, i knows about this 0-chain in r'. If $\ell = 0$, then either $(i_0, 0) \rightarrow_r (i, m)$, hence $(i_0, 0) \rightarrow_{r'} (i, m)$, or the initial value of i_0 was changed to 1 in r', so this is not in fact a 0-chain.

It remains to show that r' has no 0-chains of length greater than ℓ. Suppose to the contrary that that $i_0, \dots, i_{\ell+1}$ is a 0-chain in r' of length $\ell + 1$. If $\ell \geq 0$, then since $\mathit{latest0}_i(r, m) = \ell$, we must have $(i_{\ell+1}, \ell + 1) \not\rightarrow_{r'} (i, m)$, for otherwise we would have $(i_{\ell+1}, \ell + 1) \rightarrow_r (i, m)$ and $i_0, \dots, i_{\ell+1}$ would be a 0-chain in r of length $\ell + 1 > \mathit{latest0}_i(r, m)$, a contradiction. If $\ell = -1$ (i.e., i does not know about any 0-chains in r), suppose there is 0-chain in r'. Then there must be some agent i_0 with an initial preference of 0 in r'. We cannot have $(i_0, 0) \rightarrow_r (i, m)$, for otherwise i would know about a 0-chain in r. But then our construction guarantees that the initial preference of i_0 in r' is 1, not 0. We conclude that r' contains no 0-chains of length at least $\ell + 1$, as desired. □

LEMMA A.12. *If $m \geq 2$, $\mathcal{I}, (r, m) \models \mathit{inv}_1(i, j, k)$, and $\mathit{latest0}_j(r, m) \leq m - 3$, then there exists a run r^* such that $\mathcal{I}, (r^*, m - 1) \models \mathit{inv}_1(i', j', k')$ for some $i', j', k' \in \mathit{Agt}$, $E(r^*, m - 2) \subseteq E(r, m - 2)$, and $\mathcal{N}(r) \cup \{k\} = \mathcal{N}(r^*) \cup \{k'\}$.*

PROOF. By Lemma A.11, there exists a point $(r^\dagger, m) \sim_j (r, m)$ such that $\mathcal{N}(r) = \mathcal{N}(r^\dagger)$, the longest 0-chain in $r^\dagger$ has length at most $m - 3$, and $E(r^\dagger, m) \subseteq E(r, m)$. Moreover, since k exhibits no faulty behavior in r, the same holds in $r^\dagger$. In addition, since $\mathcal{I}, (r, m) \models \mathit{decided}_i = \bot \land \mathit{decided}_j = \bot \land \mathit{decided}_k = \bot$, we also have that $\mathcal{I}, (r^\dagger, m) \models \mathit{decided}_i = \bot \land \mathit{decided}_j = \bot \land \mathit{decided}_k = \bot$.

We construct a run r' by modifying $r^\dagger$ so that agent i is faulty, k is nonfaulty, and agent k hears from every other agent except agent i in round m. Since the only modifications are in round m, we have $E(r', m - 1) = E(r^\dagger, m - 1) \subseteq E(r, m - 1)$. Agent j is nonfaulty in both r and r' and has the same local state in both (r, m) and (r', m). Hence, $j \in (\mathcal{N} \land \mathcal{O})(r, m) \cup (\mathcal{N} \land \mathcal{O})(r', m)$; therefore, (r', m) is $(\mathcal{N} \land \mathcal{O})$-$\boxdot$-reachable from (r, m) through agent j. Moreover, the length of the longest 0-chain in r' is $m - 3$, because we do not modify $r^\dagger$ in round $m - 2$ or earlier in constructing r'. Since $m \geq 2$ and $\mathit{latest0}_k(r', m) \leq m - 3$, in r', agent k decides 1 in round $m + 1$ by Corollary A.8.

Next consider the run r'' that is identical to r' except that agent i hears from every agent in round $m - 1$. Here we have $E(r'', m - 2) = E(r', m - 2) \subseteq E(r, m - 2)$. Since the messages received by agent k are identical in round m, we have $k \in (\mathcal{N} \land \mathcal{O})(r', m) \cup (\mathcal{N} \land \mathcal{O})(r'', m)$. Thus, (r'', m) is $(\mathcal{N} \land \mathcal{O})$-$\boxdot$-reachable from (r', m) through agent k. Because the longest 0-chain in r' has length $m - 3$, and we make no change to r' in round $m - 2$ or earlier in constructing r'', the longest 0-chain in r'' aiso has length $m - 3$. By Corollary A.8, agent i decides 1 in round m of r'' upon hearing from every agent in round $m - 1$. Since agent j heard from agent i and j did not decide earlier in run r'', agent j decides 1 in round $m + 1$ in r''.

Now consider the run r^* that is identical to r'' except that agent i is in $\mathcal{N}$ (so that agent k does get a message from agent i in round m), and k is faulty in $\mathcal{N}$ (but does not exhibit any faulty behavior.) Clearly, $E(r^*, m - 2) = E(r'', m - 2) \subseteq E(r, m - 2)$. Since j has the same local state in both (r'', m) and (r^*, m), and is nonfaulty in both r'' and r^*, we have $j \in (\mathcal{N} \land \mathcal{O})(r'', m) \cup (\mathcal{N} \land \mathcal{O})(r^*, m)$. Therefore, (r^*, m) is $(\mathcal{N} \land \mathcal{O})$-$\boxdot$-reachable from (r'', m) through agent j.

Finally, observe that in run r^*, i is nonfaulty and decides 1 in round m, k is nonfaulty and does not decide before round m, and j is faulty but exhibits no faulty behavior and does not decide before round m. Hence, $\mathcal{I}, (r^*, m - 1) \models \mathit{inv}_1(k, i, j)$. By the transitivity of the $(\mathcal{N} \land \mathcal{O})$-$\boxdot$-reachability relation, $(r^*, m - 1)$ is $(\mathcal{N} \land \mathcal{O})$-$\boxdot$-reachable from (r, m), and the claim holds with $(i', j', k') = (k, i, j)$. Each step of the construction involved a swap of faultiness between a faulty and nonfaulty agent among i, j, and k, leaving the faultiness of other agents invariant, so we have $\mathcal{N}(r^*) \cup \{k'\} = \mathcal{N}(r) \cup \{k\}$. □

In the following, we write $r[0..m]$ for the prefix of the run r up to and including to time m, but removing the information about which agents are nonfaulty from the environment state, and write $Sends_i(r,m)$ for the set of agents j such that $(i, m-1) \rightarrow_r (j,m)$.

Lemma A.13. *If* $\mathcal{I},(r,m) \models inv_i(i,j,k)$ *and the longest* 0*-chain in* r *has length at most* $m-2$*, then for all* $p \notin \mathcal{N}(r) \cup \{k\}$*, there exists a* $(\mathcal{N} \wedge \mathcal{O})$*-*$\boxdot$*-reachable point* (r',m') *with* $m' \in \{m, m+1\}$ *such that* $r[0..m-1] = r'[0..m-1]$*,* $\mathcal{I},(r',m') \models inv_i(i',j',k')$*,* $p \notin \{i',j',k'\}$*, and* $(p,m-1) \not\rightarrow_{r'} (j',m')$*.*

Proof. We construct an $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$-path that establishes the result. We remark that inv_1 is generally not maintained along this path: we falsify inv_1, but re-establish it in the final step. The construction has several branches, depicted in Figure 1, in which we show just the changes made to obtain each successive run, and track failure edges in the causality graph. Agent timelines are depicted horizontally. Failure edges are indicated by dashed lines, and nodes labelled 1 indicate that the agent is deciding 1 at that node.

If $(p,m-1) \not\rightarrow_r (j,m)$, we can take $(r',m') = (r,m)$ and $(i',j',k') = (i,j,k)$, and we are done. Otherwise, consider the run r^1 obtained by modifying r so that in round m, p's messages to all agents but j and itself fail, that is, $Sends_p(r^1,m) = \{p,j\}$. We have $\mathcal{N}(r^1) = \mathcal{N}(r)$. Since $(r,m) \sim_j (r^1,m)$ and $j \in (\mathcal{N} \wedge \mathcal{O})(r,m) \cap (\mathcal{N} \wedge \mathcal{O})(r^1,m)$, we have that (r^1,m) is $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$ reachable from (r,m).

It remains the case that i, j and k exhibit no faulty behavior in r^1, and that they are undecided at time m. Since k is faulty but exhibits no faulty behavior in r^1, by Lemma A.10, no agent decides using the common knowledge conditions in r^1. It also remains the case that there is no 0-chain with length greater than $m-2$, so no agent decides 0 after round $m-1$ in r^1. If $\mathcal{I},(r^1,m) \models deciding_i = 1$, then we are done, taking $(r',m') = (r^1,m)$ and $(i',j',k') = (j,i,k)$, since $i \in \mathcal{N}(r^1)$ but $(p,m-1) \not\rightarrow_{r^1} (i,m)$ and $r[0..m-1] = r^1[0..m-1]$ by construction. On the other hand, if $\mathcal{I},(r^1,m) \models deciding_k = 1$, then we would similarly be done, taking $(r',m') = (r^{1'},m)$ and $(i',j',k') = (j,k,i)$, where $r^{1'}$ is the run obtained from r^1 by setting $\mathcal{N}(r^{1'}) = \mathcal{N}(r) \cup \{k\} \setminus \{i\}$.

If $\mathcal{I},(r^1,m) \models deciding_i = \bot \wedge deciding_k = \bot$, we have $\mathcal{I},(r^1,m+1) \models deciding_i = 1$, since $\mathcal{I},(r^1,m) \models deciding_j = 1$ and $j \in \mathcal{N}(r^1)$. Let r^2 be the run obtained by modifying r^1 so that in round $m+1$, agent i receives messages from all agents but p. We have $\mathcal{N}(r^2) = \mathcal{N}(r)$. Since $(r,m) \sim_j (r^2,m)$ and $j \in (\mathcal{N} \wedge \mathcal{O})(r,m) \cap (\mathcal{N} \wedge \mathcal{O})(r^1,m)$, we have that (r^2,m) is $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$ reachable from (r,m).

It remains the case that k is faulty but exhibits no faulty behavior in r^2, hence no agent decides using the common knowledge conditions in r^2. Note that $r^2[0..m-1] = r^1[0..m-1] = r[0,m-1]$ and $Sends_p(r^2,m) = Sends_p(r^1,m) = \{p,j\}$. We also have $r^2_i(m) = r^1_i(m)$ and $r^2_k(m) = r^1_k(m)$, so $\mathcal{I},(r^1,m) \models deciding_i = \bot \wedge deciding_k = \bot$. Thus, using Corollary A.8, we have that $\mathcal{I},(r^2,m+1) \models deciding_i = 1$, because the longest 0-chain in r^2 has length at most $m-2$.

Let r^3 be the run obtained by modifying r^2 so that $\mathcal{N}(r^3) = \mathcal{N}(r) \cup \{k\} \setminus \{j\}$, and agent k receives a message from all agents but j in round $m+1$. Since $(r^2,m)+1 \sim_i (r^3,m+1)$ and $i \in (\mathcal{N} \wedge \mathcal{O})(r^2,m+1) \cap (\mathcal{N} \wedge \mathcal{O})(r^3,m+1)$, we have that (r^3,m) is $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$ reachable from (r,m). We remark that inv_1 does not hold in r^3 with a permutation of i,j,k, since j now exhibits faulty behavior. However, we still have $\mathcal{I},(r^3,m+1) \models \neg C_{\mathcal{N}}(t\text{-}faulty)$, because $i \in \mathcal{N}(r^3)$ and i does not observe j's faulty behavior. It follows that no agent decides in r^3 using the common knowledge conditions before round $m+3$.

Note that $Sends_p(r^3,m) = Sends_p(r^2,m) \setminus \{j\} = \{p,j\}$ and $Sends_j(r^3,m+1) = Agt \setminus \{k\}$. We have $r^3[0..m-1] = r^2[0..m-1] = r[0,m-1]$. This implies that none of i,j,k has decided at time m. In addition, $r^3_i(m) = r^2_i(m)$ and $r^3_k(m) = r^2_k(m)$, so neither i nor k decides in round $m+1$. By

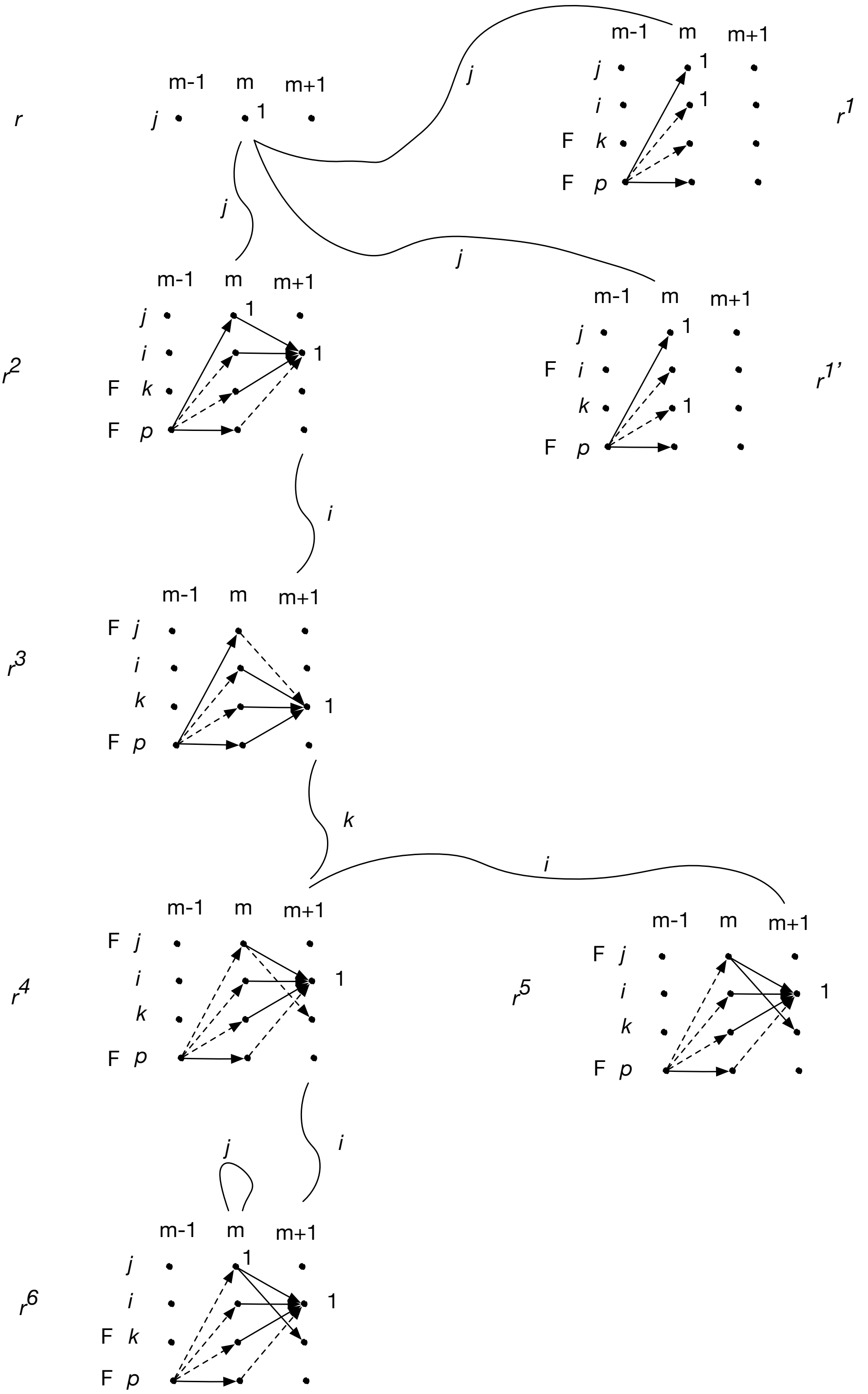

Fig. 1. Construction for Lemma A.13

Corollary A.8, we have that $\mathcal{I}, (r^3, m+1) \models \mathit{deciding}_k = 1$, because the longest 0-chain in r^3 has length at most $m - 2$.

Let r^4 be the run obtained by modifying r^3 so that agent p does not send any messages to agent j in round m, and in round $m+1$, all agents except p send a message to agent i. We have $\mathcal{N}(r^4) = \mathcal{N}(r^3) = \mathcal{N}(r) \cup \{k\} \setminus \{j\}$ and $r^4[0..m-1] = r^3[0..m-1] = r[0, m-1]$. In addition, $\mathit{Sends}_p(r^4, m) = \mathit{Sends}_p(r^3, m) \setminus \{j\} = \{p, j\} \setminus \{j\} = \{p\}$, and $\mathit{Sends}_j(r^4, m+1) = \mathit{Sends}_j(r^3, m+1) = \mathit{Agt} \setminus \{k\}$. Since in r^3, agent j does not send a message to agent k in round $m+1$, the change in round m is not visible to agent k at time $m+1$, nor is the round $m+1$ change visible to k at time $m+1$. Thus, we have $(r^3, m+1) \sim_k (r^4, m+1)$. In addition, $k \in (\mathcal{N} \wedge \mathcal{O})(r^3, m+1) \cap (\mathcal{N} \wedge \mathcal{O})(r^4, m+1)$, so $(r^4, m+1)$ is $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$ reachable from (r, m). We have $r_i^4(m) = r_i^3(m)$, so agent i does not decide in round $m+1$ in r^4. By Corollary A.8, we get that $\mathcal{I}, (r^4, m+1) \models \mathit{deciding}_i = 1$.

It is still the case in r^4 that agent j exhibits faulty behavior. To reinstate inv_1, let r^5 be the run obtained from r^4 by changing the failed message from j to k in round $m+1$ to be successfully transmitted. Since this was the only failure of j introduced earlier, this ensures that j does not exhibit faulty behavior in r^5. We have $\mathcal{N}(r^5) = \mathcal{N}(r^4) = \mathcal{N}(r) \cup \{k\} \setminus \{j\}$ and $r^4[0..m-1] = r^3[0..m-1] = r[0, m-1]$. The latter means that no agent decides 0 after round $m-1$. The change made in constructing r^5 is not visible to agent i at time $m+1$, so we have $(r^4, m+1) \sim_i (r^5, m+1)$. We have $i \in (\mathcal{N} \wedge \mathcal{O})(r^4, m) \cap (\mathcal{N} \wedge \mathcal{O})(r^5, m)$, so (r^5, m) is $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$ reachable from (r, m).

In addition, $r_i^5(m) = r_i^4(m)$, so agent i does not decide in round $m+1$ of run r^5, since it did not do so in round $m+1$ of r^4. Similarly, $r_k^5(m) = r_k^4(m)$, so agent k does not decide in round $m+1$ of run r^5. With respect to agent j, we have two possibilities.

- If $\mathcal{I}, (r^5, m) \models \mathit{deciding}_j = 1$, then let r^6 be the run identical to r^5 except that $\mathcal{N}(r^6) = \mathcal{N}(r)$. (That is, we switch k from being nonfaulty in r^5 to faulty in r^6, and j from being faulty to being nonfaulty.) We still have that $(r^4, m+1) \sim_i (r^6, m+1)$ and $i \in (\mathcal{N} \wedge \mathcal{O})(r^4, m+1) \cap (\mathcal{N} \wedge \mathcal{O})(r^6, m+1)$, so $(r^6, m+1)$ is $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$ reachable from (r, m). We also have that $\mathcal{I}, (r^6, m) \models \mathit{deciding}_j = 1$, because $r_j^6(m) = r_j^5(m)$. Since $(r^6, m) \sim_j (r^6, m)$ and $i \in (\mathcal{N} \wedge \mathcal{O})(r^6, m) \cap (\mathcal{N} \wedge \mathcal{O})(r^6, m)$, we have that (r^6, m) is $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$ reachable from (r, m). Note that $(p, m-1) \not\rightarrow_{r^6} (j, m)$. Hence we are done, taking $(r', m') = (r^6, m)$ and $(i', j', k') = (i, j, k)$.
- If $\mathcal{I}, (r^5, m) \models \neg \mathit{deciding}_j = 1$, then we have $\mathcal{I}, (r^5, m+1) \models \mathit{decided}_i = \mathit{decided}_j = \mathit{decided}_k = \bot$, and $\mathcal{I}, (r^5, m+1) \models \mathit{deciding}_i = 1$. Moreover, $\mathit{Sends}_p(r^5, m) = \{p\}$ and $i \notin \mathit{Sends}_p(r^5, m+1)$. Thus $(p, m-1) \not\rightarrow_{r^5} (i, m+1)$. Hence we are done, taking $(r', m') = (r^5, m+1)$ and $(i', j', k') = (k, i, j)$.

□

The following lemma allows us to reduce the size of the set $S_{m-1}(r)$, so that there are fewer paths by which a 0-chain reaching time $m-2$ is visible at time m.

Lemma A.14. *If $\mathcal{I}, (r, m) \models \mathit{inv}_1(i, j, k)$ and the longest 0-chain in r has length at least $m-2$, then there exists a run r' and a permutation i', j', k' of i, j, k such that $\mathcal{I}, (r', m) \models \mathit{inv}_1(i', j', k')$, $\mathcal{N}(r) \cup \{k\} = \mathcal{N}(r') \cup \{k'\}$, the point (r', m) is $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$-reachable from (r, m), and there are no 0-chains in r' of length $m-2$.*

Proof. Consider an agent $q \in \mathit{Agt}(\mathcal{N}(r) \cup \{k\})$. We write $F_q(r, m-1)$ for the set of edges of the form $(p, m-2) \rightarrow (q, m-1)$ in $E(r, m-1)$, where $p \in \mathit{Agt} \setminus \mathcal{N}(r) \cup \{k, q\}$.

We first show that there exists a point (r^q, m) that is $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$-reachable from (r, m) such that $\mathcal{I}, (r^q, m) \models \mathit{inv}_1(i^q, j^q, k^q)$, where $\{i, j, k\} = \{i^q, j^q, k^q\}$, $\mathcal{N}(r) \cup \{k\} = \mathcal{N}(r^q) \cup \{k^q\}$, and $E(r^q, m-1) \subseteq E(r, m-1) \setminus F_q(r, m-1)$. To do this, we first apply Lemma A.13 to (r, m) and

$(N \wedge \mathcal{O})$-$\boxdot$-reach a point (r^1, m_1) such that $\mathcal{I}, (r^1, m_1) \models inv_1(i_1, j_1, k_1), N(r) \cup \{k\} = \mathcal{N}(r^1) \cup \{k_1\}$, $E(r^1, m-1) = E(r, m-1)$, $(q, m-1) \not\to_{r^1} (j_1, m_1)$, and m_1 is either m or $m+1$.

Let r^2 be the run obtained from r^1 by converting all edges $(p, m-2) \to_{r^1} (q, m-1)$ in $F_q(r, m-1)$ to failures, so that $E(r^2, m-1) = E(r^1, m-1) \setminus F_q(r, m-1) = E(r, m-1) \setminus F_q(r, m-1)$. We have $(r^2, m_1) \sim_{j_1} (r^1, m_1)$, and $\mathcal{N}(r^1) = \mathcal{N}(r^2)$, so (r^2, m_1) is $(N \wedge \mathcal{O})$-$\boxdot$-reachable from (r^1, m_1). It remains the case that none of i_1, j_1, k_1 have decided at (r^2, m_1), since the state of these agents at time $m_1 - 1$ is visible to j_1 at time m_1 in r^1 and $(r^2, m_1) \sim_{j_1} (r^1, m_1)$. Hence $\mathcal{I}, (r^2, m_1) \models inv_1(i_1, j_1, k_1)$ and $\mathcal{N}(r) \cup \{k\} = \mathcal{N}(r^2 \cup \{k_1\}$.

If $m_1 = m$ we are done, taking $r^q = r^2$; otherwise, $m_1 = m+1$ and we apply Lemma A.12 to obtain a point (r^q, m) $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$ reachable from (r, m) with $\mathcal{I}, (r^q, m_1) \models inv_1(i_q, j_q, k_q)$ and $\mathcal{N}(r^q) \cup \{k^q\} = \mathcal{N}(r^2) \cup \{k_1\} = \mathcal{N}(r) \cup \{k\}$. We have $E(r^q, m-1) = E(r^q, m_1 - 2) \subseteq E(r^2, m_1 - 2) = E(r^2, m-1) = E(r, m-1) \setminus F_q(r, m-1)$.

We successively repeat the steps above for all agents in $Agt \setminus (\mathcal{N}(r) \cup \{k\})$, thereby $(N \wedge \mathcal{O})$-$\boxdot$-reaching a point (r^3, m) with $\mathcal{I}, (r^3, m) \models inv_1(i_3, j_3, k_3)$, $\mathcal{N}(r^3) \cup \{k_3\} = \mathcal{N}(r) \cup \{k\}$, and $E(r^3, m-1) \subseteq E(r, m-1) \setminus (\bigcup_{q \in Agt \setminus (\mathcal{N}(r) \cup \{k\})} F_q(r, m-1))$. That is, in round $m-1$ of r^3, no messages are transmitted between distinct (faulty) agents in $Agt \setminus (\mathcal{N}(r) \cup \{k\})$.

It may still be the case that we have faulty agents q that decide in round $m-1$ of r^3, for which $(q, m-1) \to_{r^3} (j', m)$. In this case, we again apply Lemma A.13 to (r^3, m) and $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$-reach a point (r^4, m_4) where $\mathcal{I}, (r^4, m_4) \models inv_1(i_4, j_4, k_4)$, $\{i_4, j_4, k_4\} = \{i, j, k\}$, and $\mathcal{N}(r^4) \cup \{k_4\} = \mathcal{N}(r^3) \cup \{k_3\} = \mathcal{N}(r) \cup \{k\}$, $r^4[0 \ldots m-1] = r^3[0 \ldots m-1]$, $(q, m-1) \not\to_{r^4} (j_1, m_4)$, and m_4 is either m or $m+1$. At this point we apply the construction of Lemma A.11 to produce from r^4 a run r^5 such that (r^5, m_4) is $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$-reachable from (r^4, m_4) via agent j_4, $\mathcal{I}, (r^5, m_4) \models inv_1(i_4, j_4, k_4)$, $\mathcal{N}(r^5 \cup \{k_4\} = \mathcal{N}(r^4) \cup \{k_4\} = \mathcal{N}(r) \cup \{k\}$, all 0-chains are visible to (j_4, m_4), and $E(r^5, m-1) \subseteq E(r^4, m-1) \subseteq E(r, m-1) \setminus (\bigcup_{q \in Agt \setminus (\mathcal{N}(r) \cup \{k\})} F_q(r, m-1))$.

We claim that q does not decide in round $m-1$ of r^5. Obviously, we do not have an edge from $(q, m-2)$ into $(\mathcal{N}(r) \cup \{k\}) \times \{m-1\}$ in r^4, otherwise r would have a 0-chain of length at least $m-1$. We also do not have an edge in r^4 from $(q, m-2)$ into $(\mathcal{N}(r) \cup \{k\}) \times \{m-1\}$, since all such edges have been eliminated. Thus, the only edge from $(q, m-2)$ is to $(q, m-1)$. Since $(q, m-1) \not\to_{r^4} (j_4, m_4)$, it follows that $(q, m-2) \not\to_{r^4} (j_4, m_4)$, and the construction of r^5 ensures that $\mathcal{I}, (r^5, m-2) \models \neg deciding_q = 0$.

If $m_4 = m+1$, we apply Lemma A.12 to obtain a point (r^6, m) that is $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$ reachable from (r, m) such that $E(r^6, m-1) \subseteq E(r^5, m-1)$, $\mathcal{I}, (r^6, m) \models inv_1(i_6, j_6, k_6)$, $\{i_6, j_6, k_6\} = \{i, j, k\}$, $\mathcal{N}(r^6) \cup \{k_6\} = \mathcal{N}(r) \cup \{k\}$, and q does not decide in round $m-1$. And if $m_4 = m$, we already have this conclusion with $r^6 = r^5$.

We successively apply this construction for all $q \in Agt \setminus (\mathcal{N}(r) \cup \{k\})$ until we have an $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$-reachable point (r^7, m) such that $\mathcal{I}, (r^7, m) \models inv_1(i_7, j_7, k_7)$, $\{i_7, j_7, k_7\} = \{i, j, k\}$, $\mathcal{N}(r^7) \cup \{k_7\} = \mathcal{N}(r^6) \cup \{k_6\} = \mathcal{N}(r) \cup \{k\}$, $E(r^7, m-1) \subseteq E(r^4, m-1)$, and $\mathcal{I}, (r^7, m-2) \models \neg deciding_q = 0$ for all $q \in Agt \setminus (\mathcal{N}(r) \cup \{k\})$. Since $\mathcal{I}, (r^7, m) \models inv_1(i_7, j_7, k_7)$ implies that no agent in $\mathcal{N}(r^7) \cup \{k_7\} = \mathcal{N}(r) \cup \{k\}$ decides in round $m-1$, we conclude that r^7 has no 0-chains of length $m-2$. We therefore have the result with $r' = r^7$. □

Lemma A.15. *If $\mathcal{I}, (r, m) \models inv_1(i, j, k)$, then there exists a run $\hat{r}$ such that $(\hat{r}, 0)$ is $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$-reachable from (r, m) and $\mathcal{I}, (\hat{r}, 0) \models \neg\exists 0$.*

Proof. We proceed by induction on m. Suppose that $\mathcal{I}, (r, m) \models inv_1(i, j, k)$, so that j decides 1 in round $m+1$ of r. No agent can decide 1 in round 1, so we have $m \geq 1$.

If $m = 1$, agent j can't know $\exists 0$ at time 1 right before j decides 1 in round 2, as otherwise j would have gotten a 0-chain and, by Lemma A.5, would decide 0. Hence, j considers possible the run $\hat{r}$ where no agent has an initial preference of 0 and the message pattern is identical to r. Clearly, $(\hat{r}, 1)$ is $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$-reachable from $(r, 1)$ through agent j.

If $m > 0$, suppose that $len_{)}j(r, m) = m'$. Note that j can't know about a 0-chain of length $m - 1$ or greater, as in that case j would decide 0 in round $m + 1$. If $m' < m - 2$, then we can immediately apply Lemma A.12 to $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$-reach a point $(r^*, m - 1)$ where the invariant condition for 1 holds. We can then apply the inductive hypothesis to conclude that there exists a run $\hat{r}$ such that $\mathcal{I}, (\hat{r}, 0) \models \neg\exists 0$ and $(\hat{r}, 0)$ is $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$-reachable from $(r^*, m - 1)$. By the transitivity of the $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$-reachability relation, $(\hat{r}, 0)$ is also $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$-reachable from (r, m) and the claim holds. And if $m' = m - 2$, then by Lemma A.14, there exists a point (r', m) that is $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$ reachable from (r, m) such that $\mathcal{I}, (r', m) \models inv_1(i_1, j_1, k_1)$ and all 0-chains in r' have length less than $m - 2$. The previous case applied to (r', m) then yields the result. □

Lemma A.16. *If $\mathcal{I}, (r, m) \models (i \in \mathcal{N} \wedge \bigcirc(decided_i = \bot))$, then i considers a point (r', m) possible at (r, m) where $\mathcal{I}, (r', m) \models inv_1(i, j, k)$.*

Proof. Suppose that at some point (r, m), $\mathcal{I}, (r, m) \models i \in \mathcal{N} \wedge \bigcirc(decided_i = \bot)$. We first observe that we must have $\mathcal{I}, (r, m) \models \neg C_{\mathcal{N}}(t\text{-}faulty)$, since otherwise, by Lemma A.4, we get a contradiction. To show that i considers a point (r', m) possible at (r, m) where $\mathcal{I}, (r', m) \models inv_1(i, j, k)$, we consider two cases:

- Suppose that $\mathcal{I}, (r, m) \models \neg K_i(t\text{-}faulty)$. By definition, there exists an agent k such that $\mathcal{I}, (r, m) \models \neg K_i \neg(k \in \mathcal{N}) \wedge \neg K_i \neg(k \notin \mathcal{N})$. Let r' be a run identical to r except $k \notin \mathcal{N}$, k does not exhibit any nonfaulty behavior in r', some agent $j \in \mathcal{N}$ hears from all other agents in round m, and the only 0-chains in run r' are the ones that i knows about. This means that the 0-chains in r' have length at most $m - 2$, so eventually all agents should decide 1. We claim that $(r, m) \sim_i (r', m)$. Suppose by way of contradiction that i has different local states in (r, m) and (r', m). Then either i hears in r' from an agent that didn't receive a message from k in round m of r that heard from k in round m of r' or i hears about a 0-chain in round m of r' that it did not know about at time m in r. In the first case, i would have known that k is faulty in (r, m), which is a contradiction; the second case contradicts the assumption that i doesn't know about such 0-chains in r'.

  By Corollary A.8, agent j decides 1 in round $m + 1$ upon hearing from all other agents in round m. Thus, $\mathcal{I}, (r', m) \models inv_1(i, j, k)$ and $(r, m) \sim_i (r', m)$.
- If $\mathcal{I}, (r, m) \models K_i(t\text{-}faulty)$, we make use of the observation that $\mathcal{I}, (r, m) \models \neg C_{\mathcal{N}}(t\text{-}faulty)$. By Proposition A.2, it must be the case that $\mathcal{I}, (r, m - 1) \models \neg dist_{\mathcal{N}}(t\text{-}faulty)$; equivalently, $\mathcal{I}, (r, m) \models \forall j \in \mathcal{N}(\neg K_j \neg(k \in \mathcal{N}))$ for some faulty agent k. Since $i \in \mathcal{N}$, we must have $\mathcal{I}, (r, m - 1) \models \neg K_i \neg(k \in \mathcal{N})$. Since $\mathcal{I}, (r, m) \models K_i(t\text{-}faulty)$, it must be the case that agent i learned that k is faulty in round m, either by not getting a message from k for the first time or by getting a message from some faulty agent k' that knew about agent k being faulty. Agent k' can't be a nonfaulty agent, since no nonfaulty agent knows that k is faulty at time $m - 1$. If i learned that k is faulty by not hearing from k in round m, then at (r, m), agent i considers possible the point (r', m), where in round m of r', some nonfaulty agent j received a message from all other agents. By Corollary A.8, in r', j decides 1 in round m. Thus, $\mathcal{I}, (r', m) \models inv_1(i, j, k)$, as desired.

  If i learned that k is faulty by hearing from some faulty agent k' that k is faulty, then i considers possible the run r' which is identical to r except that j and k receive a message from all other agents except k' in round m. By Corollary A.8, both agents j and k are about to decide 1 at (r', m), since all 0-chains in r' have length at most $m - 2$ and they both receive a message from all agents but k'. We must have $(r, m) \sim_i (r', m)$, as the only change between the runs is in round m, and i receives the same messages in both runs. Moreover, i is nonfaulty in both runs.

Now consider a run r'' where i does not hear from k' in round m or later, and otherwise receives all messages from the other agents. Thus, i, j, and k hear from all agents but k' in rounds m and $m+1$ in r''. As $j \in (\mathcal{N} \wedge \mathcal{O})(r'', m') \cup (\mathcal{N} \wedge \mathcal{O})(r', m')$, (r'', m') is $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$-reachable from (r', m).

Since j does not hear from k' after round m in r'', it does not discover that k is faulty. Therefore, there exists a run r''' that is identical to r' except that agent k does not exhibit nonfaulty behavior. Thus, $\mathcal{I}, (r''', m') \models \mathit{inv}_1(i, j, k)$. Moreover, $j \in (\mathcal{N} \wedge \mathcal{O})(r''', m') \cup (\mathcal{N} \wedge \mathcal{O})(r'', m')$, so (r''', m') is $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$-reachable from (r'', m').

□

Finally, we construct the points $(r_i^{0'}, m)$ and $(r^{1'}, m)$ required for weak safety. The argument for this case is simpler than that of the previous case. Most of the work for this case is done by Lemmas A.17 and A.18, which are analogues of Lemmas A.15 and A.16.

Lemma A.17. *If $\mathcal{I}, (r, m) \models \mathit{inv}_0(i, j, k)$, then there exists a run $\hat{r}$ such that $(\hat{r}, 0)$ is $(\mathcal{N} \wedge \mathcal{Z})$-$\boxdot$-reachable from (r, m) and $\mathcal{I}, (\hat{r}, 0) \models \neg\exists 1$.*

Proof. We proceed by induction on m.

If $m = 0$, then since agent j decides 0 in round 1, it must be the case that $\mathcal{I}, (r, 0) \models \mathit{init}_j = 0$. Hence, agent j considers a run $\hat{r}$ where every agent has initial preference 0 possible. Clearly, $(\hat{r}, 0)$ is $(\mathcal{N} \wedge \mathcal{Z})$-$\boxdot$-reachable from $(r, 0)$. If $m > 1$, since agent k is faulty but does not exhibit faulty behavior in r, $\mathcal{I}, (r, m) \models \neg C_{\mathcal{N}}(\textit{t-faulty})$. By Lemma A.5, agent j must have received at least one 0-chain of length m in order to decide 0. Let k' be an agent from whom j received a 0-chain. Note that k' can't be i or j, since i and j do not decide before round $m+1$. Let r' be a run identical to r, except that i is faulty instead of k, k hears from k', and k does not hear from i in round m of r'. Since $j \in (\mathcal{N} \wedge \mathcal{Z})(r, m) \cap (\mathcal{N} \wedge \mathcal{Z})(r', m)$ and j has the same local state at both points, (r', m) is $(\mathcal{N} \wedge \mathcal{Z})$-$\boxdot$-reachable from (r, m).

In r', agent k is nonfaulty and receives a 0-chain from k' in round m. Let k^* be an agent that sent a 0-chain to k' in round $m-1$. Consider a run r'' that is identical to r' except that agent i receives a 0-chain in round $m-1$ from k^* and agent j receives a message from agent i in round m. (If $m = 1$, then we take r'' to be a run where i has initial preference 0 and j receives a message from i in round 1.) Agent i decides 0 in round m of r''. Since k has the same local state in both (r', m) and (r'', m), and $k \in (\mathcal{N} \wedge \mathcal{Z})(r', m) \cap (\mathcal{N} \wedge \mathcal{Z})(r'', m)$, (r'', m) is $(\mathcal{N} \wedge \mathcal{Z})$-$\boxdot$-reachable from (r', m). Because j received a message from i in round m of r'' and j is nonfaulty, j decides 0 in round $m+1$ of r''.

Let r''' be a run that is identical to r'' except that agent i is nonfaulty. Then j has the same local state in both (r'', m) and (r''', m), and $j \in (\mathcal{N} \wedge \mathcal{Z})(r'', m) \cap (\mathcal{N} \wedge \mathcal{Z})(r''', m)$. Thus, (r''', m) is $(\mathcal{N} \wedge \mathcal{Z})$-$\boxdot$-reachable from (r, m) and in r''', agent i decides 0 in round m, $i \in \mathcal{N}$, $j \in \mathcal{N}$, $k \notin \mathcal{N}$, and k acts nonfaulty throughout run r'''. Moreover, $\mathcal{I}, (r''', m-1) \models \mathit{inv}_0(j, i, k)$. By the inductive hypothesis, it follows that there exists a run $\hat{r}$ such that $(\hat{r}, 0)$ is $(\mathcal{N} \wedge \mathcal{Z})$-$\boxdot$-reachable from $(r''', m-1)$ and $\mathcal{I}, (\hat{r}, 0) \models \neg\exists 1$. By the transitivity of the $(\mathcal{N} \wedge \mathcal{Z})$-$\boxdot$-reachability relation, $(\hat{r}, 0)$ is also $(\mathcal{N} \wedge \mathcal{Z})$-$\boxdot$-reachable from (r, m) and the claim holds. □

Lemma A.18. *If $\mathcal{I}, (r, m) \models (i \in \mathcal{N} \wedge \bigcirc(\mathit{decided}_i = \bot))$, then i considers a point (r', m) possible at (r, m) where $\mathcal{I}, (r', m) \models \mathit{inv}_0(i, j, k)$.*

Proof. Suppose that $\mathcal{I}, (r, m) \models i \in \mathcal{N} \wedge \bigcirc(\mathit{decided}_i = \bot)$. We must have $\mathcal{I}, (r, m) \models \neg C_{\mathcal{N}}(\textit{t-faulty}))$, since otherwise, by Lemma A.4, we get a contradiction. We next show that, whether or not i knows which agents are faulty, i considers a point (r', m) possible at (r, m) from which a point (r''', m) is $(\mathcal{N} \wedge \mathcal{O})$-$\boxdot$-reachable from (r, m) such that $\mathcal{I}, (r''', m) \models inv_0(i, j, k)$.

- If $\mathcal{I}, (r, m) \models \neg K_i(t\text{-}faulty)$ then, by definition, there exists a $k \in Agt$ such that $\mathcal{I}, (r, m) \models \neg K_i \neg(k \in \mathcal{N}) \wedge \neg K_i \neg(k \notin \mathcal{N})$. Let r' be a run identical to r except that $k \notin \mathcal{N}$ and k does not exhibit faulty behavior in r'. We claim that $(r, m) \sim_i (r', m)$. Suppose by way of contradiction that i has different local states in (r, m) and (r', m). Since the only difference between these runs are the blocked messages sent by k that are no longer blocked, i must have heard from an agent (either directly or indirectly) that did not receive a message from k in r but did in r'. But in that case, i would have known that k is faulty in r, which is a contradiction.
  Since i has not yet decided by (r', m) and does not decide in round m (as i has the same local state at (r, m) and (r', m)), we have that $\mathcal{I}, (r', m) \models \neg K_i(\bigwedge_{j \in Agt} \neg(deciding_j = 0))$. That is, i considers it possible that there exists an agent j' that decides 0 in round $m + 1$. Let r'' be the run that is identical to r' except that j' decides 0 in round $m + 1$ of r''. By Lemma A.5, if $m \geq 1$, j' receives a 0-chain in r'' from some agent k' in round m, and if $m = 0$, $init_{j'} = 0$. Let r''' be a run that is identical to r'' except that agent j (who is also nonfaulty) receives a message from k' in round m if $m \geq 1$ and has initial preference 0 if $m = 0$. Hence, agent j decides 0 in round $m + 1$ of r'''. By the transitivity of the knowledge relation, we then have $(r, m) \sim_i (r''', m)$.
- If $\mathcal{I}, (r, m) \models K_i(t\text{-}faulty)$, we use the observation that $\mathcal{I}, (r, m) \models \neg C_{\mathcal{N}}(t\text{-}faulty)$. As in the proof of Lemma A.16, we must have that $\mathcal{I}, (r, m - 1) \models \neg K_i \neg(k \in \mathcal{N}))$ for some nonfaulty agent k.
  It must be the case that agent i learned that k is faulty in round m, due to either not getting a message from k for the first time or getting a message from some faulty agent k' that knew that k was faulty. (Agent k' can't be nonfaulty, since no nonfaulty agent knows that k is faulty at time $m - 1$.) Since agent i has not decided yet and does not decide in round $m + 1$ of r, agent i knows that the other nonfaulty agents did not decide in an earlier round. Moreover, in (r, m), i considers a point (r', m) possible where $\mathcal{I}, (r', m) \models (deciding_{j'} = 0)$ for some agent j', since i does not decide 1 in round $m + 1$ of r'.
  If agent i learned that k is faulty due to not hearing from k in round m of r', then at the point (r', m), agent i considers (r'', m) possible, where r'' is such that some nonfaulty agent j received a message from k in round m and k decides 0 in round $m - 1$ (due to hearing from the agent that sent a message to the agent that sent a message to j' in the 0-chain). In r'', j receives a 0-chain in round m and thus decides 0 in round $m + 1$. It follows that $\mathcal{I}, (r'', m) \models inv_0(i, j, k)$.
  If agent i learned that k is faulty due to hearing it from some faulty agent k', then agent i considers a point (r'', m) possible at (r', m), where r'' is identical to r' except that in r'', agent j does not hear from agent k' at or after round m and receives a 0-chain from agent j' in round m. If $k \neq k'$, then again we have $I, (r'', m) \models inv_0(i, j, k)$.
  If $k = k'$, then $I, (r'', m) \not\models inv_0(i, j, k)$, because k does not act nonfaulty throughout run r''. In that case, consider a run r^* that is identical to r'' except that agent k sends a message to all agents but j in round m of r^*. Since $j \in (\mathcal{N} \wedge \mathcal{Z})(r'', m) \cap (\mathcal{N} \wedge \mathcal{Z})(r^*, m)$ and j has the same local state at both points, (r^*, m) is $(\mathcal{N} \wedge \mathcal{Z})$-$\boxdot$-reachable from (r'', m). Let r''' be the run that is identical to r^* except that agent k does not exhibit faulty behavior in r''' and i receives a 0-chain from j'. By construction, $\mathcal{I}, (r''', m) \models inv_0(i, j, k)$ Since agent i did not know k was faulty in r^*, $i \in (\mathcal{N} \wedge \mathcal{Z})(r^*, m) \cap (\mathcal{N} \wedge \mathcal{Z})(r''', m)$ and (r''', m) is $(\mathcal{N} \wedge \mathcal{Z})$-$\boxdot$-reachable from (r^*, m). Again, this suffices for the desired result.

□

This completes the proof that $\mathbf{P}^1$ is weakly safe and we get the following result:

Theorem 7.7. *$\mathbf{P}^1$ is weakly safe with respect to $\gamma_{\mathit{fip},n,t}$.*

A.2.7 Implementation. We now explicitly define the full-information context using communication graphs similar to those used by Moses and Tuttle [12]. Intuitively, a communication graph for agent i is a labeled graph that provides a compact description of all messages sent and received by agent i. Formally, the *communication graph $G_{i,m}$* is defined as follows. The set $V(G_{i,m})$ of vertices consists of all pairs of the form (j, m') for all $m' \leq m$ and agents j; the set $E(G_{i,m})$ of edges consists of all edges from $(j, m'-1)$ to (j', m') for $j, j' \in Agt$ and $m' \leq m$; there is a message label $l_{G_{i,m}} \in \{0, 1, ?\}$ for each edge (see below), and an initial preference label $p_{G_{i,m}} \in \{0, 1, ?\}$ for each $j \in Agt$ (which can be viewed as a label on vertices of the form $(j, 0)$). An edge from $(j, m'-1)$ to (j', m') is labeled with a 1 if i knows that j sent a message to j' in round m'; it is labeled with a 0 if i knows that j did not send a message to j' in round m'; and it is labeled with a ? if i does not know whether j sent a message to j' in round m'. Note that with a full-information protocol, if i knows the initial preferences of agents, and which agents sent round m'' messages for $m'' < m'$, then it is easy for i to figure out what the content of a message that was sent would be. A preference label of $v \in \{0, 1\}$ on $(j, 0)$ indicates that i knows that j's initial preference was v, while a label of ? indicates that i does not know j's initial preference. We write $G_{i,m}(r)$ for the communication graph of agent i at time m in a run r in a full-information exchange and $\mathcal{G}_{i,m}$ for the set of all time-m communication graphs for agent i.

Let $\mathcal{E}_{\mathit{fip}}(n)$ be the full information-exchange protocol for n agents, where for each agent i, the following hold:

- The local states have the form $\langle \mathit{time}_i, \mathit{decided}_i, \mathit{init}_i, G_{i,\mathit{time}_i} \rangle$, where G_{i,time_i} is a communication graph.
- The initial local states of each agent i have the form $\langle 0, \bot, \mathit{init}_i, G_{i,0} \rangle$, where $G_{i,0} \in \mathcal{G}_{i,0}$. (Note that that in $G_{i,0}$, we must have $p_{G_{i,0}}(j) = ?$ for all agents $j \neq i$ and $p_{G_{i,0}}(i) = \mathit{init}_i$.)
- $M_i = \mathcal{G}_{i,\mathit{time}_i}$.
- For all agents j and actions a, $\mu_{ij}(\langle \mathit{time}_i, \mathit{decided}_i, \mathit{init}_i, G_{i,\mathit{time}_i} \rangle, a) = G_{i,\mathit{time}_i}$.
- $\delta_i(\langle \mathit{time}_i, \mathit{decided}_i, \mathit{init}_i, G_{i,m} \rangle, a, (m_1, \ldots, m_n)) = \langle \mathit{time}_i + 1, \mathit{decided}_i', \mathit{init}_i, G_{i,\mathit{time}_i+1} \rangle$, where $G_{i,\mathit{time}_i+1} \in \mathcal{G}_{i,\mathit{time}_i+1}$ is obtained by adding vertices and edges for round time_i+1 and combining the labels from all graphs that were received by i and G_{i,time_i}. More precisely, if RG_i consists of all the graphs that i received up to and including round $\mathit{time}_i + 1$, then

$$l_{G_{i,\mathit{time}_i+1}}((j,m),(j',m+1)) = \begin{cases} v & \text{if } \exists G \in RG_i(l_G((j,m),(j',m+1)) = v) \land v \in \{0,1\}, \\ 1 & \text{if } m = \mathit{time}_i \land j' = i \land m_j \neq \bot, \\ 0 & \text{if } m = \mathit{time}_i \land j' = i \land m_j = \bot, \\ ? & \text{otherwise.} \end{cases}$$

$$p_{G_{i,\mathit{time}_i+1}}(j) = \begin{cases} v & \text{if } \exists G \in S, v \in \{0,1\}((p_G(j) = v)), \\ ? & \text{otherwise.} \end{cases}$$

Finally, $\mathit{decided}_i'$ is determined by the action a, just as in the standard EBA context.

Let $\gamma_{\mathit{fip},n,t} = (\mathcal{E}_{\mathit{fip}}(n), SO(t), \pi_{\mathit{fip},n})$, where $\pi_{\mathit{fip},n}$ interprets the standard propositions in the standard way. To check the knowledge conditions in $\mathbf{P}^1$, it is useful to define the following sets, which can be computed in polynomial time from the communication graph:

- For $m' < m$, $f(j, m', G_{i,m})$ is the set of faulty agents that i knows that j knows about at time m', given $G_{i,m}$. The set $f(j, m', G_{i,m})$ is the union of (a) $f(j', m'-1, G_{i,m})$ for all j' that sent a message to j in round m' in $G_{i,m}$ if $m' > 0$, (b) $\{j'\}$ for each agent j' that did not send a message to j in round m' in $G_{i,m}$, and (c) $f(j, m'-1, G_{i,m})$ if $m' > 0$. (Note that $f(j, 0, G_{i,m}) = \emptyset$.)

- For $m' < m$, $D(S, m', G_{i,m})$ is the set of faulty agents that i knows that the agents in S know about at time m', given $G_{i,m}$. $D(S, m', G_{i,m}) = \cup_{k \in S} f(k, m', G_{i,m})$.

In addition to f and D, agent i can compute the actions of each agent j at time $m' < m$ if $(j, m') \rightarrow_r (i, m)$ where r is the current run, since we are using a full-information protocol. Let $d(j, m', G_{i,m}) \in \{0, 1, \bot, ?\}$ represent what i knows about the action of agent j in round $m'+1$. If $d(j, m', G_{i,m}) = v \in \{0, 1\}$, then i knows that that j decides v in round $m'+1$; if $d(j, m', G_{i,m}) = \bot$, then i knows that j does not decide in round $m'+1$; finally, if $d(j, m', G_{i,m}) = ?$, then $(j, m') \not\rightarrow_r (i, m)$.

The set of values known by each agent that sent a message either directly or indirectly to i can be also computed from the communication graph in polynomial time. Let $V(j, m', G_{i,m})$ be the set of values that i knows that j knows about at time m' if $(j, m') \rightarrow_r (i, m)$ according to $G_{i,m}$ and $\emptyset$ otherwise.

We next define families $common_v(i, m, G_{i,m})$ and $cond_1(i, m, G_{i,m})$ of Booleans that can also be computed in polynomial time.

Definition A.19 ($common_v$ and $cond_v$). Intuitively, $common_v$ holds if $K_i(dist_{\mathcal{N}}(t\text{-}faulty) \wedge (\bigwedge_{j \in \mathcal{N}}(K_j \bigcirc no\text{-}decided_j(v))) \wedge (\bigvee_{j \in \mathcal{N}} K_j(\exists v)))$ holds at time m (which means that $C_{\mathcal{N}}(t\text{-}faulty \wedge no\text{-}decided_{\mathcal{N}}(1) \wedge \exists v)$ holds at time $m+1$), given $G_{i,m}$. We compute $common_v$ as follows. If either of the following three conditions hold, then $common_0(i, m, G_{i,m}) = false$:

- $|D(\bar{f}(i, m, G_{i,m}), m-1, G_{i,m})| \neq t$, where $\bar{f}(j, m', G_{i,m}) = Agt - f(j, m', G_{i,m})$ (these are the agents that i thinks might be nonfaulty at time m', given $G_{i,m}$);
- there exists an agent $j \notin f(i, m, G_{i,m})$ such that $d(j, m', G_{i,m}) = 1 - v$ for some $m' < m$.
- for all agents $j \notin D(\bar{f}(i, m, G_{i,m}), m-1, G_{i,m})$, $v \notin V(j, m-1, G_{i,m})$

Otherwise, $common_v(i, m, G_{i,m}) = true$.

The first condition for taking $common_v(i, m, G) = false$ corresponds to agent i thinking that the agents who might be nonfaulty at time m do not have distributed knowledge of t faulty agents at time $m-1$. If so, certainly the agents who are actually nonfaulty will not have distributed knowledge of who the faulty agents are at time $m-1$, so there will not be common knowledge among the nonfaulty agents of who the faulty agents are at time m—see Lemma 5. The second condition holds if some agent j that that i considers possibly nonfaulty at time m has decided $1-v$. We can assume that $|D(\bar{f}(i, m, G_{i,m}), m-1, G_{i,m})| = t$ (otherwise $common_v(i, m, G) = false$ by the first condition). Thus, i knows who the nonfaulty agents are at time m, so i knows that a nonfaulty agent has decided $1-v$, so it cannot be common knowledge among the nonfaulty agents that no nonfaulty agent decided $1-v$.

The Boolean $cond_0 = true$ holds if the formula $init_i = 0 \vee \bigvee_{j \in Agt} K_i(jdecided_j = 0)$ holds at time m, given $G_{i,m}$. Formally,

- $cond_0(i, 0, G_{i,0}) = (init_i = 0)$.
- For $m > 0$, $cond_0(i, m, G_{i,m}) = true$ if there exists an agent j such that $d(j, m-1, G_{i,m}) = 0$ and $l_{G_{i,m}}((j, m-1), (i, m)) = 1$; otherwise, $cond_0(i, m, G_{i,m}) = false$.

The Boolean $cond_1(i, m, G_{i,m}) = true$ holds if $K_i(\bigwedge_{j \in Agt} \neg(deciding_j = 0))$ holds at time m, given $G_{i,m}$. Formally,

- $cond_1(i, 0, G_{i,0}) = false$.
- For $m > 0$, let m' be the latest time such that $d(j, m', G_{i,m}) = 0$ for some agent j (as usual, $m' = -1$ if $d(j, m', G_{i,m}) \neq 0$ for all agents j), and let m_j be the latest time that $(j, m_j) \rightarrow_r (i, m)$, where r is a run for which $G_{i,m}$ describes i's view at time m in r. (There are many

such runs; it does not matter which one we choose, since they all agree on the whether $(j, m_j) \to_r (i, m)$.) Intuitively, $m' = len_j(r, m)$. If, for all m'' with $m' < m'' \leq m$, there exist at least $m'' - m'$ agents j such that $d(j, m'', G_{i,m}) = ?$, then $cond_1(i, m, G_{i,0}) = true$; otherwise $cond_1(i, m, G_{i,0}) = false$.

Using these definitions, we can define an implementation of $\mathbf{P}^1$ in the full-information context. Let P^{opt} be the EBA decision protocol implemented by the following program:

Program: P_i^{opt}

```
if $decided_i \neq \bot$ then noop
else if $common_0(i, time_i - 1, G_{i,time_i})$ then decide$_i(0)$
else if $common_1(i, time_i - 1, G_{i,time_i})$ then decide$_i(1)$
else if $cond_0(i, time_i, G_{i,time_i})$ then decide$_i(0)$
else if $cond_1(i, time_i, G_{i,time_i})$ then decide$_i(1)$
else noop
```

The following lemma shows that the initial conditions that are checked in the definition of $common_v$ correspond to checking for $C_{\mathcal{N}}(t\text{-}faulty)$.

Lemma A.20. *$|f(i, m, G_{i,m}(r))| = |D(\bar{f}(i, m, G_{i,m}(r)), m - 1, G_{i,m}(r))| = t$ for some agent i if and only if $\mathcal{I}_{\gamma_{fip,n,t},\mathbf{P}^1}, (r, m) \models C_{\mathcal{N}}(t\text{-}faulty)$.*

Proof. Let $\mathcal{I} = \mathcal{I}_{\gamma_{fip,n,t},\mathbf{P}^1}$ and $G = G_{i,m}(r)$. Suppose that $|f(i, m, G)| = |D(\bar{f}(i, m, G), m-1, G)| = t$ for some agent i. We first observe that $Agt - f(i, m, G) = \mathcal{N}$, as there are t faulty agents. This implies that the set $D(\bar{f}(i, m, G), m - 1, G)$ is the set of all faulty agents that are known by the nonfaulty agents at time $m-1$. Since $|D(\bar{f}(i, m, G), m-1, G)| = t$, it must be the case that for all faulty agents j, there exists a nonfaulty agent that knows that agent j is faulty. Hence, $\mathcal{I}, (r, m) \models \ominus dist_{\mathcal{N}}(t\text{-}faulty)$. By Proposition A.2, we have $\mathcal{I}, (r, m) \models C_{\mathcal{N}}(t\text{-}faulty)$.

Conversely, suppose that $\mathcal{I}, (r, m) \models C_{\mathcal{N}}(t\text{-}faulty)$. Again, by Proposition A.2, we get that $\mathcal{I}, (r, m) \models \ominus dist_{\mathcal{N}}(t\text{-}faulty)$. By definition, the union of all faulty agents known by nonfaulty agents at time $m-1$ is the set of all faulty agents. Hence, $|D(\mathcal{N}, m-1, G)| = t$. Therefore, $\mathcal{N} = \bar{f}(i, m, G)$. □

Theorem A.21. *If $n - t \geq 2$, then P^{opt} implements $\mathbf{P}^1$ in the full-information EBA context $\gamma_{fip,n,t}$.*

Proof. Let $\mathcal{I} = \mathcal{I}_{\gamma_{fip,n,t},\mathbf{P}^1}$ and $G = G_{i,m}(r)$. We show that for all points (r, m), $P_i^{opt}(r_i(m)) = (\mathbf{P}_i^1)^{\mathcal{I}}(r_i(m))$.

- If $P_i^{opt}(r_i(m)) =$ noop because $decided_i \neq \bot$ in $r_i(m)$, we clearly also have $(\mathbf{P}_i^1)^{\mathcal{I}}(r_i(m)) =$ noop, because $\mathcal{I}, (r, m) \models K_i(decided_i \neq \bot)$.
- If $P_i^{opt}(r_i(m)) =$ decide$_i(0)$ by the second line, we must have $decided_i = \bot$ in $r_i(m)$ and $m > 0$. By the definition of $common_0$, we must also have (a) $|D(\bar{f}(i, m - 1, G), m, G)| = t$, (b) $d(j, m', G) \neq 1$ for all $j \notin f(i, m, G)$ and $m' < m$, and (c) $0 \in V(j, m - 1, G)$ for some $j \notin f(i, m, G)$. From (a), it follows that $\mathcal{I}, (r, m) \models C_{\mathcal{N}}(t\text{-}faulty)$ using Lemma A.20. From (b), it follows that no nonfaulty agent decides 1 at any round $m' < m + 1$. Hence, $\mathcal{I}, (r, m) \models no\text{-}decided_{\mathcal{N}}(1)$. Finally, (c) implies that i knows that a nonfaulty agent j had an initial preference 0 at time $m - 1$; that is, $\mathcal{I}, (r, m) \models \ominus K_j(\exists 0)$. Combining these observations, using Proposition A.2, we can conclude that $\mathcal{I}, (r, m) \models C_{\mathcal{N}}(t\text{-}faulty \land no\text{-}decided_{\mathcal{N}}(1) \land \exists 0)$. By Lemma A.3, $\mathcal{I}, (r, m) \models K_i(C_{\mathcal{N}}(t\text{-}faulty \land no\text{-}decided_{\mathcal{N}}(1) \land \exists 0))$, so $(\mathbf{P}_i^1)^{\mathcal{I}}(r_i(m)) =$ decide$_i(0)$.

- If $P_i^{opt}(r_i(m)) = \mathtt{decide}_i(1)$ by the third line, we must have $decided_i = \bot$ in $r_i(m)$, $common_0(i, m, G) = false$, $common_1(i, m, G) = true$, and $m > 0$. As before, $common_1(i, m, G) = true$ implies that $\mathcal{I}, (r, m) \models C_{\mathcal{N}}(t\text{-}faulty) \land no\text{-}decided_{\mathcal{N}}(0) \land \ominus K_j(\exists 1)$, so $\mathcal{I}, (r, m) \models K_i(C_{\mathcal{N}}(t\text{-}faulty \land no\text{-}decided_{\mathcal{N}}(0) \land \exists 1))$.
  It thus suffices to show that $\mathcal{I}, (r, m) \models \neg C_{\mathcal{N}}(t\text{-}faulty \land no\text{-}decided_{\mathcal{N}}(1) \land \exists 0)$. Since $common_0(i, m, G) = false$, we must either have (a) $|D(\bar{f}(i, m-1, G), m-1, G)| \neq t$, (b) $|D(\bar{f}(i, m, G), m-1, G)| = t$ and for some $j \notin f(i, m, G)$ and $m' < m$, $d(j, m', G) = 1$, or (c) $|D(\bar{f}(i, m-1, G), m-1, G)| = t$ and for all $j \notin f(i, m, G)$, $0 \notin V(j, m-1, G)$. If (a) holds, then Lemma A.20 implies that $\mathcal{I}, (r, m) \models \neg C_{\mathcal{N}}(t\text{-}faulty)$. If (b) holds, then, as we have observed, $\mathcal{I}, (r, m) \models \neg C_{\mathcal{N}}(t\text{-}faulty)$ and, in addition, $d(j, m', G) = 1$ for some nonfaulty j and $m' < m$. Hence, $\mathcal{I}, (r, m) \models \neg no\text{-}decided_{\mathcal{N}}(1)$. If (c) holds, then we again have $\mathcal{I}, (r, m) \models \neg no\text{-}decided_{\mathcal{N}}(1)$; moreover, $0 \notin V(j, m-1, G)$ for all nonfaulty j. Hence, i considers it possible that none of the nonfaulty agents knows about a 0 at time $m - 01$ given G, so by Proposition A.2, $\mathcal{I}, (r, m) \models \neg C_{\mathcal{N}}(\exists 0)$. Thus, in all cases $\mathcal{I}, (r, m) \models \neg C_{\mathcal{N}}(t\text{-}faulty \land no\text{-}decided_{\mathcal{N}}(1) \land \exists 0)$, so $(\mathbf{P}_i^1)^{\mathcal{I}}(r_i(m)) = \mathtt{decide}_i(1)$.
- If $P_i^{opt}(r_i(m)) = \mathtt{decide}_i(0)$ by the fourth line, we must have $decided_i = \bot$ in $r_i(m)$, $common_0(i, m, G) = common_1(i, m, G) = false$, and $cond_0(i, m, G) = true$. Again, $common_0(i, m, G) = false$ implies that $\mathcal{I}, (r, m) \models \neg C_{\mathcal{N}}(t\text{-}faulty \land no\text{-}decided_{\mathcal{N}}(1) \land \exists 0)$. Similarly, $common_1(i, m, G) = false$ implies that $\mathcal{I}, (r, m) \models \neg C_{\mathcal{N}}(t\text{-}faulty \land no\text{-}decided_{\mathcal{N}}(0) \land \exists 1)$. We thus need to show only that $\mathcal{I}, (r, m) \models init_i = 0 \lor \bigvee_{j \in Agt} K_i(jdecided_j = 0)$. We proceed by induction on m. For the base case, if $cond_0(i, 0, G) = true$, by definition, it must be the case that $init_i = 0$. For the inductive step, suppose that $cond_0(i, m, G) = true$ for some $m > 0$. By definition, this implies that $d(j, m-1, G) = 0$ and $l_G((j, m-1), (i, m)) = 1$ for some agent j. Thus, $(\mathbf{P}_i^1)^{\mathcal{I}}(r_i(m)) = \mathtt{decide}_i(0)$. Then agent j decides 0 and agent i hears from agent j in round m. It follows that $\mathcal{I}, (r, m) \models K_i(jdecided_j = 0)$, so $(\mathbf{P}_i^1)^{\mathcal{I}}(r_i(m)) = \mathtt{decide}_i(0)$.
- If $P_i^{opt}(r_i(m)) = \mathtt{decide}_i(1)$ by the fifth line, we must have $decided_i = \bot$ in $r_i(m)$, $common_0(i, m, G) = common_1(i, m, G) = cond_0(i, m, G) = false$, and $cond_1(i, m, G) = true$. We also have $m > 0$ since $cond_1(i, m, G) = true$. As before, $common_0(i, m, G) = common_1(i, m, G) = false$ implies that the common knowledge conditions don't hold. We thus need to show that $\mathcal{I}, (r, m) \models \neg(init_i = 0 \lor \bigvee_{j \in Agt} K_i(jdecided_j = 0))$ and $\mathcal{I}, (r, m) \models K_i(\bigwedge_{j \in Agt} \neg(deciding_j = 0))$.
  By definition, $cond_0(i, m, G) = false$ implies that for all $j \in Agt$, $d(j, m-1, G) \neq 0$ or $l_G((j, m-1), (i, m)) \neq 1$, so either j did not decide 0 in round m or i did not receive a message from agent j in round m. In either case, $\mathcal{I}, (r, m) \models \neg K_i(jdecided_j = 0)$, so $\mathcal{I}, (r, m) \models \neg(init_i = 0 \lor \bigvee_{j \in Agt} K_i(jdecided_j = 0))$. Finally, if $cond_1(i, m, G) = true$, by Proposition A.7. we can conclude that $\mathcal{I}, (r, m) \models K_i(\bigwedge_{j \in Agt} \neg(deciding_j = 0))$, so $(\mathbf{P}_i^1)^{\mathcal{I}}(r_i(m)) = \mathtt{decide}_i(1)$.
- If $P_i^{opt}(r_i(m)) = \mathtt{noop}$ by the last line, we must have $decided_i = \bot$ in $r_i(m)$, and $common_0(i, m, G) = common_1(i, m, G) = cond_0(i, m, G) = cond_1(i, m, G) = false$. If $m = 0$, none of the conditions in $\mathbf{P}^1$ can hold except $\mathcal{I}, (r, m) \models init_i = 0 \lor \bigvee_{j \in Agt} K_i(jdecided_j = 0)$. However, since $cond_0(i, 0, G) = false$, we must have $init_i \neq 0$ and $(\mathbf{P}_i^1)^{\mathcal{I}}(r_i(0)) = \mathtt{noop}$. If $m > 0$, then arguments above show that the common knowledge conditions don't hold and $\mathcal{I}, (r, m) \models \neg(init_i = 0 \lor \bigvee_{j \in Agt} K_i(jdecided_j = 0))$. We thus need to show only that $\mathcal{I}, (r, m) \models \neg K_i(\bigwedge_{j \in Agt} \neg(deciding_j = 0))$. Since the common knowledge conditions don't hold, we can apply Proposition A.7 to conclude that this is the case. Thus, none of the conditions in $\mathbf{P}^1$ hold and we have $(\mathbf{P}_i^1)^{\mathcal{I}}(r_i(m)) = \mathtt{noop}$.

□

We can then conclude that P^{opt} is also optimal with respect to full-information exchange. Since each condition in P^{opt} can be checked in polynomial time in the size of the communication graph, and the communication graph itself uses $O(n^2 t)$ bits, we then get the following result:

Proposition 7.9. *There exists a polynomial-time implementation P^{opt} of $\mathbf{P}^1$ with respect to a full-information exchange.*

A.3 Proof for Section 8

Proposition 8.2. *If r is a failure-free run, then*

(a) If there is at least one agent with an initial preference of 0 in r, then all agents decide by round 2.
(b) If all agents have an initial preference of 1, then all agents decide by round $t + 2$ with P^{min} and by round 2 with P^{basic} and P^{fip}.

Proof. For the first part, suppose that some nonfaulty agent has an initial preference of 0. Clearly that agent decides 0 in the first round and tells all the other agents, who decide in the second round (for all three protocols).

For the second part, suppose that all the agents are nonfaulty and have an initial preference of 1. Then with P^{min}, since no agent will decide 0 or hear about a decision of 0, the agents will wait for $t + 1$ rounds of information exchange decide 1 in round $t + 2$. With P^{basic} and P^{fip}, no agent will decide right away and since all agents i will get a message from every other agent j in the first round from which they can conclude that j's initial preference was 1 ($(init, 1)$ in the case of P^{basic} and an explicit message saying that j's initial preference was 1 in the case of P^{fip}), agents can all decide on 1 in round 2. □

ACKNOWLEDGMENTS

Alpturer and Halpern were supported in part by AFOSR grant FA23862114029. Halpern was additionally supported in part by ARO grants W911NF-19-1-0217 and W911NF-22-1-0061. The Commonwealth of Australia (represented by the Defence Science and Technology Group) supported this research through a Defence Science Partnerships agreement. We thank Yoram Moses for useful comments on the paper.

REFERENCES

[1] A. Castañeda, Y. A. Gonczorowski, and Y. Moses. 2014. Unbeatable consensus. In *Proc. 28th International Conference on Distributed Computing (DISC '14)*. 91–106.

[2] A. Castañeda, Y. Moses, M. Raynal, and M. Roy. 2017. Early decision and stopping in synchronous consensus: a predicate-based guided tour. In *Proc. 5th International Conference on Networked Systems (NETYS 2017)*. 206–221.

[3] B. Coan. 1986. A communication-efficient canonical form for fault-tolerant distributed protocols. In *Proc. 5th ACM Symposium on Principles of Distributed Computing*. 63–72.

[4] C. Dwork and Y. Moses. 1990. Knowledge and common knowledge in a Byzantine environment: crash failures. *Information and Computation* 88, 2 (1990), 156–186.

[5] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi. 1995. *Reasoning About Knowledge.* MIT Press, Cambridge, MA. A slightly revised paperback version was published in 2003..

[6] J. Y. Halpern and R. Fagin. 1985. A formal model of knowledge, action, and communication in distributed systems: preliminary report. In *Proc. 4th ACM Symposium on Principles of Distributed Computing*. 224–236.

[7] J. Y. Halpern and Y. Moses. 1990. Knowledge and common knowledge in a distributed environment. *J. ACM* 37, 3 (1990), 549–587.

[8] J. Y. Halpern, Y. Moses, and O. Waarts. 2001. A characterization of eventual Byzantine agreement. *SIAM J. Comput.* 31, 3 (2001), 838–865. https://doi.org/10.1137/S0097539798340217 A preliminary version of the paper appeared in *Proc. 9th ACM Symposium on Principles of Distributed Computing*, 1990, pp. 333-346..

[9] X. Huang and R. van der Meyden. 2013. Symbolic synthesis of knowledge-based program implementations with synchronous semantics. In *Theoretical Aspects of Rationality and Knowledge: Proc. Fourteenth Conference (TARK 2013).*

[10] X. Huang and R. van der Meyden. 2014. Symbolic synthesis for epistemic specifications with observational semantics. In *Proc. 20th International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS 2014)*. 455–469.

[11] L. Lamport. 1978. Time, clocks, and the ordering of events in a distributed system. *Commun. ACM* 21, 7 (1978), 558–565.

[12] Y. Moses and M. R. Tuttle. 1988. Programming simultaneous actions using common knowledge. *Algorithmica* 3 (1988), 121–169. https://doi.org/10.1007/BF01762112

[13] G. Neiger and M. R. Tuttle. 1993. Common knowledge and consistent simultaneous coordination. *Distributed Computing* 6, 3 (1993), 334–352.