Authority (I): A Promise Theoretic Formalization

Mark Burgess

April 23, 2021

Abstract

Authority is a central concept in social systems, but it has a variety of meanings. Promise Theory offers a simple formalized understanding of authority, and its origins, as polarization within a network of collaborative interactions. This idealized approximation stands in contrast to the usual deontic view of authority in socio-philosophical literature, and unifies the various interpretations with a single idea. It's shown that the elementary meanings of authority can all be understood as a promise, analogous to that of a 'compass direction' within some decision space, with which agents may choose to align voluntarily. Authority is therefore separated from the embodiment by any particular agency or kind of agent, and is closely related to the concept of leadership in management science. Agents may try to impose authoritative directives onto subordinates, but imposition will generally be ineffective, due to their autonomy or causal independence. Stable configurations may be formed from resonant interactions that employ both semantics and dynamics to bind agents. This simple-minded formalization serves as an foundation for later study about the dynamics of authority and derived 'power'.

Contents

1	Introduction	2
1.1	Definitions from literature	2
1.2	Outline	3
2	Promise Theory and voluntary cooperation	4
2.1	Notation and conventions	5
2.2	Relevance to notions of authority	5
3	Authority as the alignment of intent either by promise or by imposition	6
3.1	Intent and the 'reference promise' as compass for alignment	7
3.2	Patterns of authority	8
3.3	Trust in authority by mandate	10
3.4	Delegation of partial authority	11
3.5	Authority as a calibrator of standards	11
3.6	Ownership as absorption of a subordinate by a superordinate	12
3.7	Direct appointment: face to face authority by invitation	13
3.8	Indirect appointment: authority imposed by third parties	15
4	Discussion of direct and indirect authority (timescales)	19
5	Scaling of authority	20
5.1	Collective authority	21
5.2	Formation of hierarchy and persistence	21
5.3	Incentives and disincentives for alignment with authority	22
5.4	Type blindness and mistaken mandates	22
6	Authority as emergent (dynamical) symmetry breaking	23
7	Summarial remarks	24

1 Introduction

The concept of authority is filled with associations and experiences from daily life. Most of us probably think we know what it is, yet the concept is loaded with subtleties from the dynamics of its origin to the semantics of its application. Our personal experiences with it likely distort our perception of its underlying nature, as a natural phenomenon—perhaps even make it seem more complicated than it is, because its effects are certainly complicated. However, in this paper, I want to show how the various interpretations of authority boil down to some quite simple Promise Theoretic patterns—rooted in the autonomy of agents and their capacity for voluntary cooperation. Following this approach, we can render authority as a phenomenon of natural science rather than humanism, loosed from spurious subjectivities and opinions about the things or persons that embody them. We might find an underlying formalization to unify all cases.

Accomplishing this, will require some forbearance from readers with strong opinions about the special nature of human phenomena. It will require some discipline around the separation of concerns and idealized approximations. In this first part, the initial step will form a foundation for the larger problem of describing the dynamics of authority on different scales, in both an informatic and socio-political setting.

1.1 Definitions from literature

In social science, conceptual definitions inevitably begin with linguistic semantics. While words are the initial anchor we have to normalize shared meaning, language can be both multivalued and ambiguous. Promise Theory offers a simple way to be precise with a formal diagrammatic and symbolic concision, but we must begin with language. A familiar search engine defines authority as:

‘The power or right to give orders, make decisions, enforce obedience’,

which places an emphasis on ‘obedience’ as a concept—i.e. a loss of autonomy. The Oxford English Dictionary, by contrast, defines authority to be

‘The power or right to influence others or act in a specified way’,

which leaves room for influence to be voluntary, without loss of autonomy. By peering through the lens of Promise Theory [1], this paper reconciles these meanings for authority with the ‘power’ and the ‘right’ to propagate influence—to invoke influence by command or by invitation. These two branches, in fact, correspond loosely to the promise theoretic terms *imposition* of intent and the *promise* of cooperation, respectively.

The literature on authority seems relatively sparse. Some recent references on the topic are helpful in reviewing and summarizing the documented philosophical semantics [2, 3]. They express authority as a pivotal concept, but keeping to the doctrine of deontic tradition prominent in philosophy and social discourse—claiming the existence of inexplicable yet inevitable forces of ‘obligation’ and ‘right’, rather than finding an interpretation in terms of autonomy, voluntary determination, and agent cooperation. This top down view suffers from the usual ‘who authorizes the authorizer?’ problem, while an autonomous agent model avoids this by providing a natural localization of authority in ‘self’.

As pointed out by Morselli and Passini [2], the word authority comes from the Latin words *auctoritas* (*auctor*), which derives from the verb ‘*augere*’ (to enhance, increase, or reinforce). *Auctor* expressed the meaning of author, creator, promoter, enhancer, which suggests the laying out of a direction or a path—an interpretation which matches well with the simple analysis in this work. Haugaard [3], on the other hand, relates the views of philosophers, summarizing three conspicuous reasons for authority from Weber’s analysis action types, where a sense of legitimacy (what I’ll call an expression of mandate in this paper) is an additional key to unlocking its meaning. What stands out from Weber is the uncommon admission of a notionally voluntary decision at play:

“Weber (1978) argued that commands of authority contain a certain minimum of voluntary submission; thus an interest (based on ulterior motives or genuine acceptance) in obedience’. This interest can either be purely practical, related to specific pragmatic interests or from the belief in its legitimacy’ (Weber, 1978: 213). The belief that sustains authority is of greater theoretical interest to Weber than crude ulterior motives as belief relates to the *Verstehen* sociological enterprise. Sociologically, permanent relations of authority are sustained by the beliefs of the grantee of authority. The various forms that the belief in legitimacy can take define the types of authority. There are three ideal types of authority which are based upon Weber’s characterization of different sources of validating legitimacy (legal rationality, value rationality, traditional action and affective action).

Weber [4] divided legitimate authority into three types: the first type discussed by Weber is legal-rational authority; the second type of authority, traditional authority, derives from long-established customs, habits and social structures; the third form of authority is charismatic authority. I believe these forms can all be accounted for using the model below.

- Rational grounds (*value*). resting on a belief in the legality of normative rules’.
- Traditional grounds (*habit or inertia to change*). resting upon established belief in the sanctity of immemorial traditions’.
- Charismatic grounds (*trust and mesmerism*). resting upon devotion to the specific and exceptional sanctity, heroism or exemplary character of an individual person ...” [3]

The italicized parentheses in the final three points have been added here as comments. No particular ranking is attached to the relative priority of these. The concept of legal authority crops up, which is a ubiquitous topic that requires explanation. In this work, I’ll show how legal convention comes about from the scaling of authority to larger groups. Of deeper interest is the admission of the role of voluntary cooperation, which aligns precisely with the Promise Theory viewpoint.

Haugaard relates further on Arendt’s philosophy—namely her remarks that on authority can be viewed as an ‘opposite’ of argumentation, violence, and coercion, in the sense that authority has to be deserved to be accepted. This underlines the primacy of autonomy again. Some of these inferences may read more into the notion of authority than is strictly necessary to explain it structurally, and the nuances might better be discussed in terms of the dynamics of ‘power’ in the human sense [5]. In Promise Theory, the remarks align with the idea that one may win a ‘mandate’ as a source of authority, i.e. where authority refers to an appointment made perhaps by group selection. Along these lines, Arendt defines overwhelming force as non-authoritative, implying that authority has to be deserved. while others might see it as a kind of ‘appointment by natural selection’.

In the light of these few sample remarks, it seems all the more important to understand the dynamical origins of authority, not merely discuss its semantics, in order to progress to an understanding that can be applied to all kinds of system whether physical, biological, social, or even informatic. By describing such a dynamical ‘physics’ for the phenomenon of authority we may offer the kind of ‘suitably idealized approximation’ to observed behaviours, which is the norm for natural science, while being fully compatible with the descriptive semantics of the interactions.

It’s striking that societies seem conditioned to think of authority through the lens of power and powerlessness, rather than as a straightforward division of roles. The former seems to belong more naturally as part of a resulting dynamical theory. The legal stability of a state is often built on the threat of overwhelming force to deter deviations from imposed ‘lawful’ behaviour, but the stability is only achieved by normalization of acceptance [5]. Only later do such behaviours become norms and habits that require only a lower level of maintenance. As a result, we perceive governments, bosses, and leaders in a historically ‘authoritarian’ light (even the word has even come to take on a pejorative meaning, versus ‘authoritative’ which is more positive). In other usage, an authority is a source of singular expertise on a particular subject.

There is understandably some cultural bias in these views. The idea of external force as the centre of explanation was also the traditional view of physics, for similar reasons [6]. The natural philosophers believed in a divine power and aligned their reasoning to be compatible with those beliefs. However, the twentieth century upturned that belief and reversed it in favour of one based on ‘locality’ and ‘autonomy’ or ‘causal independence’. Meanwhile, Promise Theory has made a convincing case for the primacy of voluntary cooperation as a phenomenon between autonomous agents, not one based on speculative coercion or invisible forces but rather on the primacy of autonomy (what modern writers call ‘agency’ in a human context). Promise Theory shows that this principle is in no way unique to qualities. This thinking has been applied to numerous scenarios now across a wide range of phenomena from information systems to sociological processes [7–11]. For some, attempting to unify the different meanings will be viewed as an anathema—for a natural scientist, the idea that a single notion can explain all cases is attractive and powerful. I ask the forbearance of the reader in this effort.

1.2 Outline

The outline of the paper is follows:

- A brief outline of Promise Theory essentials, where the starting assumption is that decisions are the autonomous capability of every agent, without the need for external authorization. This is the

assumed ground state of every individual agent, and any deviations from this condition must be explained by a voluntary choice to subordinate by aligning with an intention offered by another agent.

- A chain of dependence for authority is described and assumed to have the following form:
 1. The existence of one or more discriminating observers or processes, which point to authority.
 2. The explicit acceptance of the appointed authority by subordinate agents.
 3. A subsequent flow of authoritative decisions by the authority, to be accepted by the subordinate, and forming a relationship over time.

We discuss how the order in which these come into play may vary in different processes.

- Agents are assumed more likely to use another agent’s decisions to override their own autonomous ability to decide if they assess there to be economic value in doing so. This may change the order in which promises are made between authority and subordinates.
- Networks of promises allow for additional roles, and several ways for a mandate to be issued to delegate decisions, including: i) directly face to face, or ii) indirectly by ‘third parties’, such as an appointed board of directors. This is related to the scaling of autonomous behaviour by groups that form ‘superagents’. Promise Theory shows how different networks lead to either voluntary cooperation (by promise) or attempted inducement by imposition. Networks of interrelated agents can achieve a kind of agency of their own (called ‘superagency’ in Promise Theory). We have to bear in mind that such virtual agents can also play a role—what is sometimes called a ‘memplex’, or a collection of ideas and practices, even ceremonies, that have some inertial or confining role that has a perceived authority of its own, e.g. ‘The System’ or ‘The State’.
- Finally, scaling of authority is considered—from individual agents to collectives and even entities which present as ‘networks of interior process’ (sometimes called memplexes). The complexity and incompleteness of information available to agents in networks implies that authorizations involving third party delegation may become difficult for potential subordinates to understand and assess, especially where networks become larger and more intricate. This may have one of two effects: agents may lose trust in the appointed authorities, or they may assess promises incorrectly based on limited or approximate information leading to a degradation of the fidelity of cooperation. Agents may believe things about their own promises that are inaccurate or even erroneous. This may be of benefit or harm to the integrity of an organization’s own promises.
- A secondary aspect of scaling is the alignment of norms amongst the population of agents. Once a norm becomes dominant, agents may perceive this as a threat to making a different choice, and thus an imposed authority becomes harder to refute. Agents will therefore be more likely to align their voluntary acceptance of an imposed decision with those of others for fear of negative consequences. This is a source of stability, for better or for worse.

For reasons of limitations on space, readers are assumed to know a minimum about the motivations and assumptions for Promise Theory (see [1] and online resources). This introduction cannot be an exhaustive look at the dynamics of authority, in relation to power. The goal here is to provide a clear and inscrutable definition on which to build in subsequent work.

2 Promise Theory and voluntary cooperation

Promise Theory is a model of autonomous cooperation expressed between generalized ‘agents’. Agents can be human, animal, vegetable, mineral, or virtual (software) in nature. For some, describing human agents on a par with non-sentient processes might be considered a kind of philosophical or even religious heresy—as Darwin experienced—but that is indeed the goal of this work, which turns out to be powerful.

Using Promise Theory, we may explore universal aspects of authority, i.e. those which do not depend specifically on the nature of agents, except in the sense of what they claim, embody, or represent—whether by original ideation or by proxy. All human lives are enveloped ‘systems’, cultural and utilitarian (memplexes), as well as tools that are not human, even when they are the products of human activity. These automate and stand in place of human behaviour by proxy. Promise Theory’s focus is on how underlying possibilities—‘intentions’ and ‘behaviours’—can be defined. These are separable and may by

carried or wielded any intermediary, thus forming a kind of ‘directionality’ that agents can subsequently align with. Promise Theory offers a description of such an impartial network of processes, which may then be scaled from an individual to mixed groups of different size.

We shall assume that the agents in a collaborative process are *a priori* independent or autonomous, meaning that their behaviours are self-determined. One could say that the unadorned ‘ground state’ of agents is one of causal independence, without bias, labels or unignorable influences. What happens to constrain those basic behaviours then depends on the voluntary interactions within groups.

2.1 Notation and conventions

The basic tenet of autonomy is that ‘no agent can make a promise about any agent other than itself’. Any agent can make decisions, but not every agent would be able to keep a promise without violating this fundamental tenet. In order to authorize access (i.e. to promise permission to access a resource one controls), the resource has to be part of (or be owned by) the promiser. Anyone can try to impose a decision, but that would require a voluntary subordination as described in this paper: only the owner of the resource can make a keepable promise about it.

A brief summary of notation is in order [1]. We write a promise about b from an agent A_1 to A_2 using an arrow

$$A_1 \xrightarrow{b} A_2. \quad (1)$$

The next point is important to those new to Promise Theory. We are not dealing with a theory of obligations in which agents necessarily must do this or that, rather we have a theory of autonomous choices. From the assumption of autonomy, an agent cannot make a promise on behalf of any agent other than itself, so this arrow implies a constraint on A_1 , not on A_2 . In order for influence to pass between them, A_2 must accept the promise to make use of it. We use the (+) sign for offer and the (-) sign for acceptance by convention:

$$A_1 \xrightarrow{+b} A_2 \quad (2)$$

$$A_2 \xrightarrow{-b} A_1. \quad (3)$$

A third kind of interaction is an ‘imposition’, written with a fist-like arrow:

$$A_1 \xrightarrow{b} \blacksquare A_2, \quad (4)$$

which is understood to mean an attempt to induce A_2 to change its behaviour, e.g. by throwing a ball and shouting ‘catch!’. The fundamental autonomy of the agents implies that this will only succeed if A_2 has already promised to accept such impositions. So the promise of (3) is needed for this too. Impositions are generally ineffective, and rely on a substrate of promises or voluntary cooperation to succeed. Readers are referred to [1] for more details.

2.2 Relevance to notions of authority

A theory of essentially voluntary cooperation, Promise Theory, may sound initially contrary to the notion of authority, but this is not the case. While the traditional view of authority is based on the notion that we are compelled either by duty or rationality to follow orders provided by some privileged body, Promise Theory proposes instead that such behaviour is emergent from within rather than induced from without. Even if one takes the view that there is no circumstance in modern society in which humans are free of the need for authorization—that we are born into circumstances beyond our control—Promise Theory explains how that view is not representative of the fundamental nature of agents and is not irreversible. It is probably not true of individuals’ activities in a hunter-gathering society (though these often have intransigent belief systems on a different level), nor would it be true of the birds in a flock, the cells in foetus, or the cars in traffic. But, surely humans are totally different from these absurd examples!? In fact, we can discuss all these on comparable terms, without prejudice. This leads to how we may understand the scaling of intent—starting from an individual’s autonomous deliberations, to such pervasive power structures as nation states, or rain forests. We expose the answer as the scaling of intent [12].

The terms and nomenclature of authority have a few related cases and meanings, worth noting for clarity:

- An *authority* may be a *thing* (an agent): appointed or self-declared to have the right to originate some kind of service on behalf of others, such as decision-making, e.g. an authority on Shakespeare or a tax authority.
- A condition of *authority* refers to the notional right (state of cooperation), embodied by the agent in the previous point, to exercise the aforementioned service e.g. possessing the authority to lead.
- An *authoritarian* agent (regime or individual) is one that imposes the aforementioned service role onto others, without seeking a mandate to legitimize the role.
- To *authorize* is to originate a (usually written) work, or to originate the service, such as decision-making.
- An *authorization* is either the act (verb) or outcome (noun) of authorizing an outcome relating to the aforementioned service.
- An agent may be assessed *authoritative* on some matter if it is accepted as an authority by others.

All of these meanings can be accounted for comfortably within the simple-minded framework of Promise Theory. Promise Theory tells us that agents are always free to act independently and voluntarily. Nevertheless, Promise Theory also grants us a model in which to compare those assumptions and what they say about the participating agents within a system. Thus, we do not begin with the assumption of omnipresent top-down authority, where one cannot make progress without seeking permission from some installed ruler, rather we begin from the bottom up as society must have done in an evolutionary picture.

It's not the purpose of this article to quarrel with any of the interpretations, rather to examine how they can be supported and explained by the *dynamics of agent interactions*. In fact, all of them seem to be straightforwardly explicable in terms of *alignment*. Rather, what Promise Theory has taught us, from experiences in human-technical systems, is that semantics emerge from the possible dynamics of a system in a way that has to be consistent with those constraints. A system can only do what it can do. An interpretation can therefore only be built on a stable behavioural knowledge, which is typically built over multiple learning interactions. So, we now turn to the interaction dynamics to look for a deeper foundation.

The making of decisions can be represented as a service abstraction. We view an individual or organization as a generic agent that can promise decisions or any other process outcome as a service:

$$A \xrightarrow{+\text{decision outcome}} A'. \quad (5)$$

The decision has to be accepted in order to be useful,

$$A' \xrightarrow{-\text{decision outcome}} A. \quad (6)$$

By the rules of Promise Theory, both these declarations are needed for autonomous intentions, expressed as promises, to lead to a propagation of the intent behind the decisions. This simple observation is all we need to know to build a consistent picture of authority from the bottom up. When we subscribe to such decisions either briefly or over time, we ascribe a sense of authority to the promise of decisions, which then typically attaches to the agent itself. This authority can be ephemeral or lasting.

3 Authority as the alignment of intent either by promise or by imposition

Promise Theory tells us that authority over other agents is never an inherent 'right'—i.e. something to impose upon them. The principle of autonomy tells us that agents themselves are the only ones who can grant that right directly. An agent's rights extend only over what resources they consist of, or have been promised by others. Subordinates may voluntarily signal their 'willingness' to accept influence, such as commands (as in military operations) or they can promise the de facto willingness to accept others' authority by promising a mandate for another to act in that role—a mandate in this sense is a kind of invitation to a position of authority. This is the meaning of 'appointment' in Promise Theory.

3.1 Intent and the ‘reference promise’ as compass for alignment

A key concept we’ll refer to in the promise theoretic understanding of authority is that of a ‘direction’ of ‘alignment’, which can be thought of as the selection of a path within a space of possible decisions. This is a simple notion of *intent* (yet to be parameterized and formalized). Again, this is a topic which has been discussed in philosophy [13–15]

A promise, in whatever language it’s expressed, is simply a declaration of such an intent in such a way as to describe the distinction between a successful outcome and an unsuccessful outcome—either the promise was kept, partially kept, or not kept. We can think of this in a fully general sense. This is what we mean by ‘direction’ and ‘measure’ in this context. For the sake simplicity we may assume that every agent has only a single intention in this discussion. This is simplistic, but it’s a useful formality.

The selection of a decision outcome from a body of possibility takes the form of a ‘service’ provided by some agent, which we typically refer to as a leader or decision-maker. Again, there is no need to think too narrowly in terms of human roles. A leader or authority could be anything, from compass pointing North or a signpost pointing the way to Rome, to a Google search that selects best hits on a search topic. This is the virtue of a model with universal characteristics. We may then say that an agent is authoritative in its promise if other agents accept it as a reliable source of the reference promise, and assess its outcomes to be of high quality (giving their trust). Notice that the exact nature of the promise is not important. The concept of authorization is closely associated with trust [16], as well as semantics like expertise, and quality—but the precise specification remains true whatever the nature of the promise offered by an agent.

Assumption 1 (Authorization) *Authorization is not a property of an agent, but is concerned with a specific promise, such as a promise to lead, to make decision, to provide information, etc. We can call this the reference promise:*

$$\text{SOURCE} \xrightarrow{+\text{reference offer}} \text{SEEKER}. \quad (7)$$

The reference promise implies a role, which is distinct from the agent that promises the role at any given moment, and may outlive any single agent (e.g. the office of the president). Authority is not conferred on the source agent unless the recipient or seeker agent accepts this promise:

$$\text{SOURCE} \xrightarrow{-\text{reference acceptance}} \text{SEEKERSOURCE}. \quad (8)$$

The degree of acceptance may be full or partial. This is a necessary condition for authorization. The resulting binding leads to a degree of authorization lies in the overlap between what is offered and what is accepted:

$$\text{degree of authorization} := \text{reference offer} \cap \text{reference acceptance}. \quad (9)$$

This overlap has the basic characteristics of a mathematical ‘inner product’, which may therefore be formalized to play the role of a degree of alignment with the quasi-direction of the reference promise within a ‘space’ of possible outcomes.

Corollary 1 (Authority) *An agent which has been authorized may be called an authority. The degree of authorization be be called the degree of authority and an individual assessment of the quality and reliability with which the agent keeps its promised role, to be made by each and every agent.*

Note, once again, for absolute clarification that an agent need not be a person, nor even an individual of any kind. It could be a collective, a government, a system of historical and cultural practices (‘memeplex’), etc, that floats on top of any number of individuals human or otherwise, as a persistent source of intentional guidance—whether in the form of books, traditions, or any other persistent memory. An agent assessed to be an authority may be called an authoritative ‘provider’ or ‘source’ of information (e.g. in technology the DNS master) is also the ultimate arbiter of the information. Similarly an expert is considered authoritative if in perceived command of the information, usually by first hand experience.

Authority can be about the right to make decisions, but it’s broader than that. Authorized agents, authorized electricians and plumbers are agents who wear a badge of membership in a certain organization which rubber stamps their promises with a certain standard promise. The conclusion, in simple dynamic terms, is that authorization is in the same class of phenomena as a kind of trust signal: a badge of membership, signalling trust in a custodian of a specific resource, which in turn aligns with the need of

agents willing to subordinate to its promise. The reference promise may be about information, work, procedural standards, or something else.

Promise Theory is quite useful in stripping away the fogginess of these concepts to reveal this simple core. The unifying concept of authority is ultimately the appointment of ‘trusted’ agents. By implication, an authoritative agent is one that promises to serve as a calibrator concerning the definition of a kind of the reference promise. We thus may define authority more carefully and generically as follows:

Definition 1 (Authority about X) *An agent S may be called authoritative by another agent R if it is the source of a promise $+X_S$,*

$$S \xrightarrow{+X_S} R, \quad (10)$$

which is fully or partially accepted by other agents,

$$R \xrightarrow{-X_R} S, \quad X_S \cap X_R \neq \emptyset. \quad (11)$$

and who in turn assess that its promised outcomes in the matter of X_S are of high value

$$\alpha_R(X_S \cap X_R) \gg 0. \quad (12)$$

In other words, such an agent is a trusted source of X to the receiving agents.

And as a corollary:

Definition 2 (Subordinates with respect to X) *The agent(s) R , who promise $-X_R$ in some degree, where the promise overlaps at least partially with the offer $+X_S$ have—in so promising—voluntarily subordinated themselves to S in the matter of X (only).*

The apparent circularity of the definition reveals that authority is not an absolute property—it’s a relativistic concept. As we see below, it makes sense as self-consistent role within a network of promises between a group of trusted agents. The role of trust may imply the need for extended or long term interactions between agents, in general, wherein they assess each other’s reliability (see [16]).

3.2 Patterns of authority

The structure of authority, within a network, can be subtle, but we can unravel it straightforwardly, in stages, using Promise Theory. The language of promises may be unfamiliar, so readers will hopefully indulge some repetition of the essential points, in what follows, for clarity. The right to determine and provide the reference promise may be granted within a network of complicated relationships, with the side effect that the weary but voluntary acceptance of imposed authority could be interpreted as an effective mandate (it masquerades as the same formal role, however illegitimate). Thus, while the ‘right’ to issue commands and directives originates entirely autonomously in every agent, any such promise or imposition must ultimately accepted by the agents who would become subordinates to the authoritative source. There are plenty of avenues for deception, incentive, disincentive, etc to shape the outcome. However, before getting into such complexities, it’s important to start, in the traditions of science, by sketching out the notions with suitably idealized approximations and simplifications that reveal the essence of what’s going on.

The interpretations of authority, as power or right, may be sketched as follows:

- *Imposition:* an authority attempts to impose its influence on a subordinate:

$$\text{Authority} \xrightarrow{+\text{influence}} \text{Subordinate}. \quad (13)$$

Because the imposee is an autonomous agent, it needs to accept the uninvited imposition to make it effective:

$$\text{Subordinate} \xrightarrow{-\text{influence}} \text{Authority}. \quad (14)$$

Without this explicit acceptance, an imposition (any attempt to induce cooperation uninvited in another agent) is powerless in the face of their autonomy.

In more complex scenarios, while direct imposition is ineffective, agents could be ‘tricked’ into acceptance of them (see section 5). If subordinates do not accept the imposition, one can still

imagine that the imposer could assert its authority by force. For instance, if the imposer could conquer, overwhelm and subsume the subordinate, then the agent would become a part of it, and autonomy would then rule over it too. By absorbing the agent within a larger superagent boundary, its autonomy is lost as far as exterior agents are concerned.

- *Promise*: the promise of authority can be accepted purely on a voluntary basis, without obligation, especially where the authoritative candidate has been volunteered a mandate M by a number of willing subordinates, to issue its commands or make decisions X :

$$\text{Subordinate} \xrightarrow{+M} \text{Authority} \quad (15)$$

$$\text{Authority} \xrightarrow{-M} \text{Subordinate} \quad (16)$$

$$\text{Authority} \xrightarrow{+X \mid M} \text{Subordinate} \quad (17)$$

$$\text{Subordinate} \xrightarrow{-X} \text{Authority}. \quad (18)$$

The unconditional promise of $+M$ is the promise theoretic understanding of an invitation by the subordinate [1]. In this way, authority is a symbiotic relationship. Followers promise their support, the authority accepts that and uses it as a basis for making singular decisions conditionally on the mandate, which are then accepted by the followers. This is part of the basis structure used in democracy. The mandate M may be interpreted as the ‘right’ or permission to lead [1].

The implications of these simple patterns are profound: unless a boss or manager’s commands can be upheld by overwhelming brute force, or perhaps by threat of later consequences (e.g. suspending wages), then he or she or their office has to seek this cooperative mandate to play its role as manager. If the manager does not actually control the policy about wages, then it does not have that leverage, and needs to maintain support by mutual cooperation. A manager or authority is therefore a *role by appointment*. The manager and subjects are in fact *coupled* in a symbiotic state of mutual subordination, rather than a unidirectional hierarchy of subordination. We can summarize these observations in two points as follows:

Definition 3 (Implicitly appointed authority) *An agent A , which is in receipt of one or more promises to subordinate themselves from agents $\{S\}$, by accepting a reference promise X , given a mandate M :*

$$A \xrightarrow{+X} \{S\} \quad (19)$$

$$\{S\} \xrightarrow{-X} A. \quad (20)$$

The common promise of $-X$ forms the appointment to the position of manager or boss. This appointment to take on the capability of deciding P is what makes the manager an ‘authority’. The ‘right to manage’ is the promise to accept P .

In this most basic form of authority, agents accept a direction $+X$ an agent, effectively appointing it to the role of leader, or trusted authority. Although this simplest form of alignment is not precisely what we typically consider to be authority in a society, it’s a pattern which is so widespread that the implications are wide, deep, and worthy of consideration by anyone: simply accepting goods or services from some agent gives that agent temporary authority to define their content. Having made this observation, we can construct a more recognizable dynamic for selection of a leadership role.

Definition 4 (Explicitly appointed authority (by invitation)) *An agent A , which is in receipt of one or more promises to subordinate themselves from agents $\{S\}$, by accepting a reference promise X , given a mandate M :*

$$\{S\} \xrightarrow{+M} A \quad (21)$$

$$A \xrightarrow{-M} \{S\} \quad (22)$$

$$A \xrightarrow{+X \mid M} \{S\} \quad (23)$$

$$\{S\} \xrightarrow{-X} A. \quad (24)$$

The common promise of $-X$ forms the appointment to the position of manager or boss. This appointment to take on the capability of deciding P is what makes the manager an ‘authority’. The ‘right to manage’ is the promise to accept P .

Notice that, in the case where agents are interacting directly ‘face to face’ and vote for their leader by promising $+M$, the acceptance of a mandate from the subordinates themselves has the implication that the appointed agent A is also subordinate to the group $\$S\}$, by a different promise, leading to a symbiotic state of mutually self-mandated authority. The direct interaction is the only case where this is true.

The broader mechanisms by which the foregoing promise exchanges arise need not be speculated on here. We treat them in the mathematical spirit of ‘necessary and sufficient conditions’ for direct authority. Perhaps, in general, a coherent balanced symbiosis might have to be seeded by the imposition of force in order to stabilize a behaviour into a habit initially, which has interesting implications for statecraft and popular moral positions on leadership. This is a matter for empirical studies to reveal and confirm or refute this theoretical picture.

3.3 Trust in authority by mandate

From the patterns above, we find two main cases to discuss, concerning the appointment of authority. Starting from the assumption of agent autonomy, it follows that agents never cede control of their decisions or behaviour under any circumstances. This leaves only two choices by which to exert authority over them:

- *Direct appointment* as in section 3.7: To undertake an asymmetric relationship voluntarily in which one is offered a mandate to be dominant over some aspects of behaviour X , or,
- *Indirect appointment* as in section 3.8: For one party to overwhelm the other so that it becomes only a part of a larger agent with autonomy.

In the latter case, one agent can effectively gobble up another (as in a corporate takeover) so that its own autonomy becomes an insignificant part of the behaviour of the conquering agent, i.e. its underlying autonomy is masked by other agents. This typically happens at scale within a collective, where the interior details need further explanation. We can sketch these scenarios roughly as in figure 1.

Figure 1: Two approaches to the power to influence. In (a) agents align voluntarily with a service. The appointment of the source of the service is by the offer of a mandate or tenure as a voluntary act. This leads to a symbiosis of mutual value. In (b), agents are conquered or subsumed by another superagent which then owns the boundary for outward expression around them. If there is any deviation of what the subordinate and conqueror voluntarily promise, the subordinate’s view may be shielded or overridden by overwhelming force.

The latter act of overwhelming another agent in order to take ownership is, in fact, a simplified rendition of a more complication process based on the former exchange of promises [9]. Using Promise Theory, we can unpick these issues to any level of detail, but no further details will be give here. The simple conclusion is that, ultimately, all authority can be seen to arise by some form of ‘voluntary’ subordination, motivated by value judgements of the agents concerned. Herein lies a connection to two person (non)-cooperative Game Theory [17, 18], in which the initial symmetry between agents is broken by ‘winning’.

3.4 Delegation of partial authority

Developing the notion of conditionality further, we can extend the semantics in a simple way to include partial authority delegated by a manager—say to middle managers or delegates D , by conditionally promising to accept a restricted mandate $M(X_D)$ to follow their decisions on a subset $X_D \in X$.

$$A \xrightarrow{+M(X_D \subseteq X)} D \quad (25)$$

$$D \xrightarrow{+M(X_D \subseteq X)} A. \quad (26)$$

Technically, this is a basic application of the matroid calibration pattern from [1]. The managers now have a mandate to decide matters $X \subseteq X_D$:

$$D \xrightarrow{+X_D | M(X_D)} S, \quad (27)$$

We are now in the realm of a new kind of mandate, which is not face to face between authority and subordinates, but involves a third party. There are further implications here for the autonomy of the agents, which we can expand on in 3.8.

There are many examples in systems where one agent or process grants the authority to act as a proxy on behalf of another agent involves intermediate agents. This introduces new issues (see the intermediate agent law in [1]). Intermediates, or ‘middle-men’, lead to great uncertainty, because as autonomous agents themselves they may not align their intentions with direct owner.

Example 1 (Power of attorney) *Granting someone to act on behalf of oneself.*

Example 2 (Payment Authorization and Proxy Resource Authorizations) *Banks can delegate access to accounts by giving a mandate to a particular credit card or payment method to represent that account, in online transactions. This is an indirect authorization. A different piece of plastic, or a note with the account number details might not be authorized.*

Authorization of any shared resource (e.g. access to a building) involves either a direct promise of access rights by the owner of the resource, or appointing a decision maker to approve access on its behalf. The decision maker has to own or control the resource at least by proxy, else it ‘doesn’t have the authority’. So actually, authority to decide is a form of ownership, or boundary management.

Example 3 (Authorized Delivery Agent) *The end to end delivery problem [1] is an example of this. A retailer promises its goods via a third party delivery agent, such as the postal service. The delivery agent is authorized to deliver packages on behalf of the retailer and the customer by a series of promises, but how do we know that the delivery agent will keep its promises exactly as the retailer and the customer intend? Of course, we don’t know—indeed, no agent knows this with certainty. The agents use trust to seal their commitment to the promises.*

3.5 Authority as a calibrator of standards

When a source is said to be authoritative, it has an implicit power (used heuristically here, in the absence of a proper definition), in the sense that it will often be accepted as the final word on the matter of X , whatever that is. It may be prioritized over other alternative sources. This power is thus over the subordinates who accept its reference promise of X (whatever that might be). The appointed role does not make the agent unique, as other authorities may have a different version of what appears to be the same promise, and may have been appointed by other agents. So the appointment of authority may have a fracturing effect causing populations of subordinates to cleave into different subsets.

Corollary 2 (Calibration) *The subordinates, as recipients of an authoritative promise, can always assesses whether two promises are equivalent and whether the outcomes between different authorities are compatible or not. An authority is thus a calibrating agent, or the arbiter of a standard, in the language of [1].*

Example 4 (Judges and courts) *Judges are indirectly appointed authorities who calibrate the interpretation of the law, comparing the interpretations of events by parties for prosecution and defence. A single judge can be trivially authoritative. A panel of judges may still disagree about its interpretation. Each judge is an authority, and the panel of judges can form its own authority as a ‘supreme’ superagent whose collective assessments are combined by from their individual assessments.*

3.6 Ownership as absorption of a subordinate by a superordinate

Ownership is related to the notion of boundaries and what they mean for agents' autonomy. By taking ownership of an agent that can be owned, the owner effectively envelops and overwhelms the other, and it becomes part of an extended superagent that contains the property of the owner. This process has been described in more detail in the context of Promise Theory and voluntary cooperation in [9]. It's not the intention to repeat that discussion here, only to point out the relevance of boundaries to what resources an agent can make promises about.

Ownership plays a role in authority because the owner of a resource is assumed to 'speak for', i.e. to literally or figuratively control it. This means that agents owned by an agent are effectively a part of it, and the agent can make promises on behalf of those and only those things it contains and owns. The use of quotes seems necessary to emphasize that we should not take words too literally in whatever direct meanings the reader would normally understand. Rather, be aware that concepts of being a part of may be physical or virtual, especially in the modern world of information. We understand a person to be in control of their possessions, at least on some level, if only to speak with some authority on their behalf—yet, they may not have direct literal control over a house or a business as if it were a part of their bodies. These details can simply be subsumed into the notion of a promise as an expression of intent rather than absolute guaranteed outcome.

Figure 2: Alignment of intent, with the subordination of some agents, can take several forms, all expressible in Promise Theory by means of promises and impositions. These illustrations capture the subtle distinctions. In (a), an agent attempts to impose its service onto a subordinate, which may or may not promise to accept the imposition. In (b), the two agents are voluntarily and accidentally aligned in their promises and therefore appear indistinguishable in their behaviours. In (c), one agent (the superordinate) absorbs the subordinate and speaks in its behalf, effectively imposing its behaviour onto the subordinate ad hoc. Note that, Promise Theory tells us not to think of this absorption as being necessarily physical—a company can take over another without physically enveloping its headquarters. Finally, in (d), two agents join into a single molecular configuration voluntarily. The superordinate promises its definition of some behaviour X and this is accepted and parroted by the subordinate voluntarily.

The conquest of an agent can be understood and interpreted in several forms, all expressible in Promise Theory by means of promises and impositions (see figure 2). Once conquered, an agent is effectively owned, hence the relationship between these terms. All of these must (in Promise Theory)

amount to a voluntary (self-authorized) promise by an agent to align itself with an outside direction.

3.7 Direct appointment: face to face authority by invitation

The basic conception of authority arises in interactions such as the two person game [17, 18]. It is a direct face to face configuration of agent promises, whose aim is to engineer a stable on-going interaction between two kinds of agent, both of whom assess beneficial value from the interaction. The result is an equilibrium configuration in which the symmetry between the agents is broken by a separation of concerns. One party effectively subordinates itself to the other with respect to one role, while the other party subordinates to the former in a different sense. We can call the two agent roles *leader* and *subordinate*. The subordinate will be the agent which makes a promise $-L$ to accept a service $+L$ from an agent in the role of leader.

The appointment of the leader is equivalent to the promise of a mandate. This is an act of invitation, as described in [1, 19]. We can summarize this direct ‘face-to-face’ configuration in words before formalizing in promises:

1. A group of voluntary subordinates A_S invites the service $+X$ of a leader agent or body A_L , effectively promising it a mandate M to lead the group. The details of the mandate need not be defined in detail, and could be quite variable—written in different languages, with different phrasing etc. What matters is only whether the mandate is accepted in the spirit in which it was intended, as this is what will lead to *alignment* with X and a stable binding.
2. If A_L accepts this invitation M , it then promises to lead (as a kind of service offering), based on this (which we write $L|M$). Again, the precise nature of what is promised is not important; what matters is whether the promise is accepted in the mutual interaction and leads to a stable separation of roles and a cohesive binding.
3. Thus, the final step is the A_S to accept the offer of leadership as promised.

Notice how all promises are made voluntarily, and the suggestion of a mandate is in no way an obligation that binds the leader to make a particular promise. The mandate and the offer of leadership based on the mandate are still independent and autonomous choices.

We can write the picture more formally in Promise Theory, in terms of two generic agents A_L and A_S representing a leader role and a subordinate role [1]. We needn’t make any assumptions about the interior structure of these agents, e.g. are they individuals or groups working together. All that matters is the promises made at the boundaries of the agents.

The agents $A_i = \{A_L, A_S\}$ are initially symmetrical, and break their symmetry (autonomously or voluntarily) when A_S appoints A_L to a special role, such as in an election or job assignment, by promising a mandate M to play that role:

$$\pi_M^{(+)} : A_S \xrightarrow{+M} A_L. \quad (28)$$

By making this promise, the symmetry between the two agent roles has already been broken to any observer (in scope) to whom the promise is observable. If A_L accepts this promised appointment,

$$\pi_M^{(-)} : A_L \xrightarrow{-M} A_S, \quad (29)$$

then there is now an axial symmetry or cephalization of the agents with a head A_L and tail A_S . Each agent can already assess the value of these promises to itself:

$$\alpha_S(\pi_M^{(+)}) = v_S(\pi_M^{(+)}) \quad (30)$$

$$\alpha_S(\pi_M^{(-)}) = v_S(\pi_M^{(-)}) \quad (31)$$

$$\alpha_L(\pi_M^{(+)}) = v_S(\pi_M^{(+)}) \quad (32)$$

$$\alpha_S(\pi_M^{(-)}) = v_S(\pi_M^{(-)}), \quad (33)$$

where $\alpha_A(\pi)$ is the assessment, by agent A of the promise π , in the usual way [1]. Each agent makes its own relative assessment, using its own units of measurement or criteria. They may or may not be based on real numbers or a discrete menu of outcomes. When matters to the collaboration is whether the agents make an assessment which leads to the continuation or stability of the promise binding.

The leader agent now makes a conditional promise to deliver its service conditional on the mandate, which is the only acceptable response to this offer of the mandate $\pi_M^{(+)}$ to A_S :

$$\pi_L^{(+)} : A_L \xrightarrow{+L | M} A_S. \quad (34)$$

Notice the conditional notation, represented by the vertical bar: a promise of leadership L ‘if I am in possession of and have accepted a kept promise of’ M . The recipients must similarly accept this offer of ‘leadership’ L to complete the binding.

$$\pi_L^{(-)} : A_S \xrightarrow{-L} A_L. \quad (35)$$

Although we call the promise L ‘leadership’, it could be any service provided by A_L . The structure of the pattern will always allow it to be identified as a form of leadership. This tells us that leadership may not be explicit, with uniforms and trappings of power, but something more subtle which is based entirely on information asymmetry.

Figure 3: Basic authorization, step-by-step, as a symmetry breaking phenomenon in timelike stages. One set of agents subordinates themselves to another by appointing a leader. Appointment occurs by promising a mandate for conditional authority to lead (appointing ‘decision rights’), which ‘authorizes’ (i) a leadership role. Assuming the mandate is accepted (ii), the recipient of the mandate can conditionally promise leadership $L|M$ (iii). If that promise is accepted, then the binding is complete (iv), and the pair of roles orients with a head and a tail in some semantic interpretation. The sustainability of this interaction configuration depends on the assessment of value over a timescale. Notice that, in this version, it’s A_L that we associate with possession of ‘authority’.

The agents can, once again, assess these new promises

$$\alpha_S(\pi_L^{(+)}) = v_S(\pi_L^{(+)}) \quad (36)$$

$$\alpha_S(\pi_L^{(-)}) = v_S(\pi_L^{(-)}) \quad (37)$$

$$\alpha_L(\pi_L^{(+)}) = v_S(\pi_L^{(+)}) \quad (38)$$

$$\alpha_S(\pi_L^{(-)}) = v_S(\pi_L^{(-)}), \quad (39)$$

and if the sum of these, for each agent, passes a certain acceptable threshold (measured in its own units), then the agents will continue to keep their promises.

$$v_S(\pi_M^{(+)}) + v_S(\pi_M^{(-)}) + v_S(\pi_L^{(+)}) + v_S(\pi_L^{(-)}) \geq V_S \quad (40)$$

$$v_L(\pi_M^{(+)}) + v_L(\pi_M^{(-)}) + v_L(\pi_L^{(+)}) + v_L(\pi_L^{(-)}) \geq V_L. \quad (41)$$

Notice that the assessment of value, a process undertaken by an agent, implies the existence of a timescale for sampling and interaction of events (messages or information) that can be associated with observation of promise keeping. The language of those messages need not be decided here. The cumulative ‘payoff’ from these individual valuations ultimately determines the course of the interaction as a two-person game [1], without needing to assume any universal scheme of value. Valuations are entirely private assessments.

Notice also that we’re assuming that the important quantity is the sum of the assessments, as one would consider in iterative Game Theory, where payoffs are real-valued quantities, accumulated over many plays or ‘moves’, but even this is not really necessary. Qualitative assessments work just as well. If each of the promise assessments passes some simple threshold of classification as ‘valuable’ that may be sufficient to settle the outcome, e.g. as a tuple:

$$(v_S(\pi_M^{(+)}), v_S(\pi_M^{(-)}), v_S(\pi_L^{(+)}), v_S(\pi_L^{(-)})) \geq (a, b, c, d). \quad (42)$$

for some constants a, b, c, d on the interior of A_S .

We interpret this basic face to face scenario as the basis of authorization. In practice, it’s a symmetry breaking of roles (see section 6). Time plays a role in the sense of partial ordering of the agents in which of them offers their side of the binding promises first, and it plays an only going role as a scale for sampling one another’s outcomes to assess (e.g. verify or validate) the truth of the compliance with the promises made. The initial assessment of value, associated with promise keeping, is a prototype for a notion of trust between agents [16]. If trust is favourable then the agents will continue to make these promises and the game will iterative in a stable manner, as predicted by [20, 21].

There is no imposition and no force involved between the agents that compels them to accede to a separation of roles. The principle of autonomy of agents is upheld. Each agent voluntary promises to adopt a certain role, even though this may be interpreted as a subordination of one by the other. The reason is that the leadership agent A_L is assessed to offer value in this role, which presumably exceeds what of the uncooperative symmetrical state. This is the phenomenon of *specialization*. Moreover, since the ‘decisions’ to accept these offered promises are made entirely autonomously by each side, we can claim that the promise of specialization leads to an emergent symmetry breaking (a cephalization or the organism in the biological sense) [12, 22, 23].

To summarize, agents that are able to promise asymmetric qualities to one another, with offer and acceptance affinities can separate into leader and subordinate. This leads to a direction which is the basis of a hierarchy. However, the concept of authorization here is in fact mutual. Without the initial mandate, the leader is not empowered to promise its leadership service, and the subordinates will not accept it. Clearly this resembles the nature of voting in a kind of democratic process; however, we should be careful in applying such a simple model. As we’ll see below, this simple mutuality does not easily survive the scaling of the process through indirection.

3.8 Indirect appointment: authority imposed by third parties

The direct face to face scenario above reveals that authority begins as a form of symbiosis. This status as a mutually beneficial interaction becomes harder to maintain as the number of agents grows, and multiple roles emerge. This balanced hierarchy is not the form of authority that we are most familiar

with in society; that's because modern societies have layers of sophistication based on separations into many roles, which place value on different promises and make their independent assessments. So we need to extend the pattern of authority by mandate include further asymmetries.

Figure 4: A common case in business and elsewhere is that leaders (e.g. CEO) are appointed by a third party, such as the board of directors of a company rather than the agents who are the recipients of the services they provide. Unlike the previous case of figure 3, it's now the third party promiser A_D of the mandate whom we associate with the having of authority or decision rights concerning the appointment of a leader, and the A_L who has decision rights over A_S . So we have a chain of authority, or a hierarchy.

Consider the scenario in figure 4. Here, the mandate for authority is no longer provided by the subordinate recipient A_S of the leadership service L , but by a third party A_D , which we could interpret as a higher level ‘director’. This is a common scenario in larger firms, for example, where a board of directors appoints a leader (giving their mandate to lead) and the leader then promises (or even imposes) their decisions based on this mandate to a different agent A_S , who has now offered no mandate for providing the service.

If we, once again, associate subordination with the voluntary act of accepting a service (leadership) $-L$. The leader agent A_L is now also subordinate to the director A_D , and the A_S are subordinate to A_L , which suggests that there might be an implicit subordination of A_S to A_D too, by transitivity. However, promises are not generally transitive, and (as yet) no direct promise has been made between A_D and A_S , so the situation lacks the necessary information to make that assessments.

Let's once again summarize in words before attaching symbols to the interactions.

1. The director offers a mandate for an agent to lead a group of (now possibly involuntary) subordinates, thus appointing a leader. The status of the subordinates is different now, because the group has not explicitly signalled its invitation to be led. The decision to appoint a leader is a kind of imposition onto the subordinates. It's not quite making a promise on behalf of another agent, so the tenets of autonomous promising are not violated, but the presence of the intermediary A_L means that the invitation to lead is not made by the potential recipients of that leadership, so the authority of the appointment is now in question.
2. Assuming that the leader accepts the mandate of the director, it promises both the director and the now ‘appointed subordinates’ its offer of leadership according to that mandate. Both of these have to accept that promise in order to create a influential binding. The value of the binding can be assessed by each. Presumably the director sees the value of this promise, having made the invitation, and would therefore accept—but the value to the subordinates is unclear. Their acceptance is less clear. It must be taken on its own merits—with a higher degree of ad hoc trust, as the subordinates have not signalled their wishes and thus the alignment of the leadership with their wishes is no longer clear by design. It may or may not align with their wishes, and thus the subordinates may or may not accept the leadership.
3. The subordinates may therefore accept or not.

The situation is now unclear. We can ask ‘on what authority does the director appoint the leader on behalf of others?’. A common and perhaps most meaningful answer is that the director is the de facto ‘owner’ of the organization, in total, and therefore controls them without question. This is a common and convenient legal fiction that builds on a litany of conventional promises. From the viewpoint of Promise Theory, it needs to be unravelled. For machinery or any other non-human agents, ownership is uncontroversial; however, the idea that humans can be fully owned is no longer widely accepted in the modern world. For example, this may mean the subordinates have already promised to be a part of the organization, in certain roles, by signing a contract (voluntarily). More usually, we consider that the owner of the organization can ‘own’ a work processes, even a slice of an agent's time, but not the human agents themselves—but that is only a convention.

Let's examine the necessary and sufficient conditions to sustain a stable configuration of agents in the Promise Theory formulation.

Figure 5: Full acceptance of a third party authorization mandate. Unlike the previous case of figure 3, it's now the giver of the mandate whom we associate with the having of authority or decision rights concerning the appointment of a leader.

In the figure, all the promises go in the same direction. However, they have neither been accepted nor do they have any causal implications until accepted. If we add acceptance promises, ad hoc, then according to the laws of assisted promising, we have (see figure 5):

$$A_D \xrightarrow{+M} A_L \quad (43)$$

$$A_D \xleftarrow{-M} A_L \quad (44)$$

$$A_L \xrightarrow{+L|M} A_S \quad (45)$$

$$A_L \xleftarrow{-M} A_S \quad (46)$$

$$A_S \xleftarrow{-L} A_L. \quad (47)$$

Notice that (because the source of $+M$ is now dissociated from A_S , A_L has to promise A_S explicitly that it has acquired and accepted such a promise to validate its conditional promise).

The natural question to ask now is why would A_S accept the promise of leadership from A_L without having chosen the agent itself, or mandated its promise of L ? The simple answer has to be that A_S still assesses a positive value to accepting this state of affairs. However, the scenario is now different because A_S no longer has a veto on that role in the same way, by being able to withdraw its mandate. At best it can refuse to accept the promise by revoking its acceptance $-L$.

The difference now is that A_S had no say in the selection of A_L . That may or may not be a problem. A_S can still vote with its feet, in a sense, and not accept A_L 's promises. As long as it does accept (see (47)), then it has effectively selected A_L . So all that remains to discuss is the stability of that configuration. What if A_S fails to assess a positive value from the service $+L|M$?

Indirection adds complications. A priori, A_S may not know about the agent A_D and its role, but it can infer that such an agent exists (in principle) by the promise (46) which indicates that A_L received its conditional mandate from a third party. If the mandate had come from A_L directly, it would have promised $+M$ as well as $+L|M$, effectively cancelling the reference to M . The fact that it promises $-M$ offers distinct information about indirection. By exposing this information to A_S , A_S may form an assessment of that information, and thus its sampling and assessment, followed by interior decisions to continue accepting the promise binding, involves sufficient information to revoke the acceptance based on whatever reasoning is implied by this assessment. If the agents are subatomic particles, we assume the reasoning is rather mechanical; if the agents are humans the reasoning could be complex.

The position of A_D in this new hierarchy seems ad hoc, in the current configuration. It may be supported by happenstance or by assessment of functional benefit, but this is not the state of affairs in organizations of agents on a more sophisticated level. For instance, in a firm A_D represents the owner of the firm. The concept of ownership does not exist for primitive collaborations, but it does exist in commerce. We can cement that authority, representing ownership as a right to decide by adding more promises—promises that we perhaps leave implicit on a day to day basis, but which form the foundation of legal ownership and acceptance of that system.

The implications of ownership are not defined clearly in general, but may be considerable. If we expand the ownership promise (figure 7) to be a promise that offers membership with associated benefits, subject to a set of conditions, then the conditions are now used as a form of incentive or coercion.

Naturally, a potential employee does not *have* to accept—they can walk away. However, this is a clear attempt to induce a promise, with a potential reward for doing so. That makes an *imposition* [1] masquerading as a bundle of promises—one might even say that involves deception, though the practice

Figure 6: Adding the concept of authority to give the mandate by prior ownership of the collaboration. In the lower figure we can suppress the other internal promises and focus on the main alignment of the collaboration, induced by the promises called ‘ownership’.

of imposing such restrictions is so widespread as to be almost common knowledge. We should be careful not to assume that common practice implies inevitability. There are many alternatives to this practice.

In spite of convention, humans may be deceived by contractual terms, which persuade unsophisticated human workers into believing that they have been bought and must accept the appointed leader out of a sense of obligation (i.e. a belief that they have promised this). This is a form of systemic authority, in which the meta authority of the system or ‘memeplex’ overwhelms or subordinates by habit. More emancipated agents, however, might take a different view. Ultimately, autonomous agents will always assess these relationships individually. The concept of ownership is complex in the modern world, and we won’t delve into the extensive infrastructure of promises it depends on here. Suffice it to say that the owner of an ownable thing is taken to control it in a meaningful sense.

Example 5 (Ownership and human rights) *The difference here is that a director (presumed owner) owns the rights to the organization resources, and acts as the voice of the entire superagent. In the case of the brain speaking for a whole animal, this is more straightforward. We tend to get confused when individual wants and desires—indeed ‘human rights’ granted by even greater agents of collective society’ that may be considered to override local considerations. The idea that human rights are intrinsic to all humans is mere rhetoric, however desirable. I’ll return to this in section 5.*

Example 6 (Democratic governance) *When an owner or director imposes itself a priori, it has its autonomous power to control (see [9]). The situation can be different in the case of indirect authorization, if the ‘owner’ or director were in fact invited into the role i.e. in possession of a voted mandate to take on the appointed role to decide. This would cement a cohesive mandate around the otherwise separated agents (see figure 6) for the entire collaboration from director, through intermediate leader to subordinates. That is the approximate state of affairs in so-called democratic governance, for example. In that case, a mandate for top level authority may be won over an extended ‘face to face’ relationship between subordinates and director (e.g. president figurehead or governing party campaigning directly to its ‘base’ of voters, thus bypassing the delegated middle layer(s)).*

The indirection described in this section proves important as a way of scaling delegation across a coherent organization, but it’s a non-trivial matter. The more layers of indirection encountered in a chain of this kind, the more likely the support for accepting and keeping the binding promises is fragile and will fail to cohere. An intermediate agent is an obvious fracture point.

Notice that, if A_S accepts the bundle of promises called ‘own’, then embedded within these is an authorization pattern, which provides the mandate to A_D to make decisions on behalf of A_S , and which

Figure 7: Promising another agent membership subject to conditions is one way to say of claiming a form of limited ownership over that agent. Here A_H promises (+) that it is the owner of A_S and A_L , and the agents accept this (-). In these imposed conditions is the expectation that A_S promise to give its mandate for authority to the agent appointed by A_D . Since this was not an autonomous choice, there is a concealed imposition here. However, impositions are known to be ineffective in general. When agents don't give a mandate freely, the acceptance of authority is indirect and the value associated with it may be considerably reduced.

may therefore include decisions about the choice of A_L :

$$A_S \xrightarrow{+M=conditions} A_D \quad (48)$$

$$A_D \xrightarrow{+L=membership|M=conditions} A_S, \quad (49)$$

as well as the matching acceptance promises. In other words, the acceptance of the imposition to promise the conditions of membership acts as a mandate to grant decision-making power to the director role through membership. The use of a conditional is interesting because it allows A_D to 'trick' A_S into promising to give a mandate for authority to A_L , by burying intentions in the fine-print. However, this is not a direct promise to A_L , so there is only a second order 'promise to make another promise to someone else' which is implicit in the relationship. This might be enough to induce the cooperation, but now the direct value of doing so is less obvious because the indirectness could make the assessment of whether that promise is kept less frequent and less rigorous. There is effectively a trust in a collaborative outcome, perhaps with a promise of penalties, etc.

The offer of employment to A_S and A_L by A_D could include a number of conditions (imposed, since they were not invited or requested). Acceptance of this bundle of promises could therefore be considered as a mandate to appoint a leader. Since the promise is to A_D , the acceptance of ownership can't include acceptance of promises made by A_L , but it would include acceptance of the decisions of A_D , which could include delegation of leadership to A_L implicitly.

In practice, these delegations are not promised explicitly, only implicitly in a model of hierarchy, which is institutionalized as a norm to become 'organizational culture'. It's sometimes referred to as learned helplessness [19].

4 Discussion of direct and indirect authority (timescales)

It's useful to compare two other scenarios of these types with slightly different structure: i) the election of a government by a vote (where the mandate to lead the voters is given by the vote) and ii) the appointment of a board of directors A_D by shareholders S of a company (where the shareholder voters are not the employees, and their mandate is once again given to a board of directors who give the mandate to a leader to lead). There are multiple levels of indirection in these chains of authorization, from face to face symbiosis to multiple levels of indirection (see figure 8).

A dynamical issue, which we haven't mentioned thus far, is that of timescales: the timescales of the promise-keeping processes are very different in different organizational structures. This affects assessments of promise-keeping as well as trust. A company might be a more agile organization than a government body, which makes decisions on a timescale of days or weeks, whereas a government makes decisions on a timescale of months and years. The assessment of outcomes, the reevaluation of trust, and the blindness to promise specifics all become coarser and less reliable with increased time—principally because the

timescale of human life-processes is fixed. Humans are unable to adapt to perceive processes on shorter or longer timescales without technological assistance.

Figure 8: Comparing two common chains of authorization with only the main promises drawn. In (a) a government election where the voters are the ones delegating authority to promise a number of laws as a trusted service, which voters promise to accept. This is an example of the case in section 3.7, so in this idealized representation a government has the direct mandate of the people to lead. Next, in (b) a share company where shareholders select a board of directors who in turn select a leader (CEO), who is delegated the authority to make decisions for the company. The principal difference between these is that i) is a closed loop, whereas ii) is an open chain of imposition in which shareholders make no promises directly to workers. This is what we often refer to as ‘authoritarian’ governance.

So, if we compare two well known chains of authorization: the election of a government, and a shareholder vote, we see a few differences. In a government election, the voters are the ones delegating authority to promise a number of laws as a trusted service, which voters promise to accept. This is an example of the case in section 3.7, so in this idealized representation a government has the direct mandate of the people to lead. In a share company, shareholders select a board of directors who in turn select a leader (CEO), who is delegated the authority to make decisions for the company. The principal difference between these is that i) is a closed loop, whereas ii) is an open chain of imposition in which shareholders make no promises directly to workers. This is what we often refer to as ‘authoritarian’ governance. In society, most workers experience the latter kind of governance on a daily basis, even when they participate in the former election process, because the effects of their voting act only on a very slow timescale.

Shareholders are self-appointed, when they are offered a mandate to purchase shares in a company. They require no further authorization to make their decisions. Nevertheless, we begin to see the potential complexity of the interactions: a company issues shares, voluntarily subordinating itself to a mixture of authoritative agents who become partial owners. They do this in return for money to sustain the company, generating a small ecosystem.

A representation of ownership in terms of voluntary cooperation might feel unfamiliar, as legal matters are typically represented as legal matters, i.e. deontic impositions by the power of the state onto its subjects. In reality, the situation is that the powerful state is only a background which documents norms and conventions associated with voluntary behaviours.

5 Scaling of authority

In physics, scaling is about how properties change or are preserved as the key sizes of system variables change. For example, a rectangle cannot be the same shape if we change only the length of one side; to preserve the shape, we have to scale width and height in proportion. So, shape has a specific scaling law based on the ratio of sides (like a television format 4:3 or 16:9). Similarly, there are different ways in which the scaling from a small number of agents to large numbers affects the derived properties of a collection of agents. Promise Theory allows us to formalize these notions from the details of promise bodies X . Promise Theory extends the notion of scaling to include semantics— not only quantitative but also qualitative measures. For example a promise might or might not depend on the nature of the agents, i.e. whether agents are human, machine, animal, vegetable, mineral, etc. There are many issues to tackle around scaling. Let’s sketch just a few suggestive cases.

5.1 Collective authority

When agents come together to behave as a single superagent (see figure 9), the behaviour of individual agents may be fully or only partially aligned, yet their dominant behaviour can still characterize the collective. On the other hand it may not—the agent could split into two.

An example of a superagent might be a company, with all of its bylaws and traditions, or a country with cultural and legal norms, but also active services that corral members into certain behaviours implicitly. Member agents might have to actively abstain from these default rails in order to reject their promises. Nevertheless, they can always do so one way or another.

In a collective, what can begin with a clear expression of a promise may evolve into a distribution of promises, qualitatively or quantitatively different—some of which may be more aligned than others. There is no need to consider a collection of agents to be a single cohesive swarm (or mob) aligned on all issues: a more reasonable picture is that a collection of agents that fall into different groups on different issues. The fact that some dominate in number might not be the decisive measure of their impact, because the latter depends not only on what is offered, but also how it is received. The seed of authority behind the ‘dominant’ directions could be different for each distinct promise type. This is consistent with the modern idea of leadership being a distributed quality based on expertise in teams and organizations, rather than being a single appointment for all issues in a single entity. It’s how we organize specialist bodies within a society, like health organizations, economic forums, environmental groups, etc, though it’s still less common in smaller groups perhaps due to resource constraints.

Figure 9: An agent that takes possession of things may (for some purposes) regard those things as part of its interior resources. Thus the superagent formed from the owner and its possessions forms a virtual superagent, which behaves for some intents and purposes as if it were an singular autonomous entity.

Of course, it’s clear from Promise Theory, that agency is not a purely individual quality. It too can be scaled. A strongly aligned group of agents can maintain a strictly consistent direction for alignment to guide a large number of subordinates authoritatively. A key mechanism here is hierarchical scaling.

5.2 Formation of hierarchy and persistence

At different scales most systems and organizations adopt a hierarchical structure. Agents form superagents, superagents form larger superagents, and so on. At each scale there can be authoritative coordination of direction from a single source, rather than a consensus approach to decision-making. This is simply explained on an economic basis [24, 25] by the fact that promises are not transactional information, they have on-going costs: a 1 to N relationship ($O(N)$) is considerably cheaper to maintain than an N to N ($O(N^2)$) relationship, as long as $N > 1$. This explains why group cohesion tends to favour singular authority structures over consensus structures, even when they are the result of a mandate by voting.

This point has obvious repercussions for societies in which agents are dynamical populations, which come and go by birth and death. When a new agent is born into a pre-existing ecosystem of promises (‘memplex’), it may experience these as ‘boundary conditions’ or ‘exterior fixtures’ rather than matters to negotiate. A social system of pre-aligned promises can survive longer than any individual agent, so what happens at society scale is not necessarily aligned with what individual agents would select had they started from scratch. In a system of voluntary cooperation, the ground state of autonomy means that agents are not compelled (except by indirect incentive) to agree to the network of promises instigated by prior circumstances and agents. Thus suppose, having established a voluntary hierarchy of authority in a social collective by voting, a child is born into such a scheme, having played no role in the selection process by which the rules and norms of society were decided. Does the authority vested by the prior population oblige the child to agree and be cooperative? The Promise Theory answer is

clearly no, though by the same token it is a matter for individual assessment—however, the penalty for non-acceptance could be to lose other rights granted by that system, which are effectively ‘owned’ by the pre-existing representative authorities. The temporal persistence of collective promises is therefore a form of collective inertia present in all multi-agent systems. There is indeed power in this simple form of ‘mob rule’ by incentive and disincentive at any scale.

Dunbar and coworkers have pointed out that cognitive limits on agents lead to a natural limit on the scaling of relationship hierarchy [26, 27]. Moreover, the intensity of the relationship (timescale for interactions) matters too—the more intense a relationship, the fewer an agent can manage. The numbers for human cognition start at 2-5 others for intense interactions, and fall into groups of 10-15 for larger teams, 30-40 for tribal groups, and 100-200 for agents whose characteristics we ‘know’.

These limitations must apply to knowledge of things as well as persons. So we can predict that authority could be represented by the placement of something as neutral as a road (e.g. the old native trail of Broadway, which defies its otherwise rectilinear promises). Our working relationships and interactions with tools consume even more of our time than our friendships in the modern world, so it’s natural that those promises are persistent. Moreover, we are bombarded with ephemeral promise offers from all quarters on different issues. As the capacity to comprehend specific promises and types fades or is coarse grained away by sheer number, cooperation may be either assisted or arrested by such lack of capacity to retain informational fidelity.

5.3 Incentives and disincentives for alignment with authority

Elementary low level agents, such as molecules and machinery, tend to have fixed and predictable behaviours. Higher level, more sophisticated agents are less predictable, as they contain many more variables and may calculate their behaviours in realtime, embody strategies to optimize their responses, etc. They can experience effective incentives or disincentives, which may shape their responses. Studying this issue in depth is a subject for a different paper, but we can point out a few examples here for future work.

The scaling of influence, when authority is diluted amongst multiple leadership appointments, always leaves the recipient (downstream) with a role in selecting between the different promises. In the case of company shareholders, there isn’t much discrimination between different mandates and conditions (but there is some, with preferred shares etc). If one takes a view based on the primacy of obligations, then it’s hard to escape from the moral trap of needing approval for one’s actions. More entrepreneurial spirits may care less for these indirect constraints. These are issues that can be studied and compared to empirical results in future work.

The threat of negative consequences could lead an agent (or superagent) to not accept a promise of directed leadership. Scaling plays a role here, because the magnitude of a threat may increase or scale with the number of agents making it. As the disparity between the value of offer and (non)acceptance grows, with number of individuals in each superagent, the assessments about repercussions for non-acceptance of an imposed direction may also grow beyond a threshold for tolerance. This is the meaning of ‘force’ in a scheme of voluntary cooperation.

Promise Theory predicts that force is not the classically deterministic force as physicists and social physicists often choose as their model [6], but rather an individual likelihood for interaction outcome, more like the mechanics of the quantum and statistical realms. Force is nonetheless still a matter for voluntary cooperation, which depends on a threshold for inertial intransigence amongst agents potentially subordinate to the authoritative force. Although every agent is free, in principle, to choose their own course, they may assess the repercussions of that choice to be unacceptable. This is especially true when agents of comparable importance come together as a ‘mob’, with sufficient number as to be potentially threatening to the well-being an agent choosing its course.

Promises to sanction agents, should they not act in a way aligned with another agent’s intentions, are how we understand threats [8]. Promise Theory predicts that threats that are imposed may or may not be effective. Promise Theory tells us that this is an autonomous assessment, for each agent individually. It depends on the extent to which the agents know and trust one another. An agent which frequently bluffs or cries ‘Wolf!’ may not be heeded and be ineffective, while an agent whose reliability in is judged to be high might be feared more. In the same way, imposition of threats (without an established trust relationship to rely on) would also be ineffective in general.

5.4 Type blindness and mistaken mandates

A limitation of scaling, in the face of limited processing capacity, is how it can affect the fidelity or integrity of promise information. The inability to process interactions in detail, at scale, may involve

the need to ‘coarse grain’ or ‘average out’ information. This much affect the resolution of information available to processes at each scale [28, 29]. Every agent’s facility for processing information is limited by its internal resources, so the reliability and timeliness of assessments made by agents falls into question at scale.

In order to keep up with cognitive load, an agent may have to sacrifice its attention to detail, i.e. its fidelity in preserving the precise essence of what is promised. This may lead to semantic errors of interpretation and associated misunderstandings. In this section, let’s consider a possibility inherent in autonomous agents of limited fidelity. As the mismatch of scales between information and resources becomes more debilitating, agents must become somewhat blind to the details of the promises they have given. It’s possible for agents to assess their own promise networks incorrectly.

This brings us back to the matter of indirect authority in section 3.8. It becomes easier to impose authority (without an accurate assessment of a mandate) at scale, because agents may not have the capacity to validate it—especially where agents are forced to overlook detail in order to preserve their normal functions. Consider, for example, the case where an agent assesses that it has, in fact, voted for a mandated leader (promised its support), when in fact it has only signed up to be a member of the organization to do a specific task, because it misunderstood the terms. This is likely a common situation in employment agreements. Employers and service providers can impose the suggestion of conditions on employment, in the small print, to ‘trick’ employees into this de facto promise unknowingly, by implication, which the co-signers of the agreements unwittingly attach their promise of cooperation to. Many legal cases have been fought in this way too, leading some countries to forbid weak contract behaviour in basic law.

Again, this phenomenon is not a human issue, but one of information resolution that occurs in any channel over which information is passed [28, 29]. The promises may seem to play similar roles from a semantic or structural perspective, connecting the same agents with different intent, but do not explicitly grant them any rights to make decisions on their behalf. This is the risk in placing interpreted meanings over and above highly specific and ‘inevitable’ dynamics. This kind of indirection may have either positive or negative consequences. It can clearly lead to great dissatisfaction if agents discover that they have agreed to be a part of an organization in which they have no decision rights—no say in decisions that affect them. On the other hand, they might accidentally be happy with the decisions being made, in which case there would be no need to rock the boat and withdraw cooperation because of formalities. This might then be called ‘emergent satisfaction’.

6 Authority as emergent (dynamical) symmetry breaking

To crown the argument for authority as a universal phenomenon, in a dynamical sense, let’s briefly divert into the fully abstract to point out that authority is part of a wider class of phenomena in nature, which exists even on an elementary level. Authority need not be treated as an exclusively human characteristic. In physics, the concept of symmetry breaking refers to a pervasive phenomenon, whereby previously indistinguishable phenomena become distinguishable, i.e. they lose their commonality or symmetry. This is true when a plain bar of metal becomes a magnet and takes on a magnetic North and South pole, or a single cell fetus develops a head and a tail. It’s also true when a vanilla population of humans divides itself up as supporters of particular sports teams or by job specialization.

Authority is an example of axial symmetry breaking, whereby a body of agents forms source and receiver ends (heads or tails) from what begins as a symmetrical state [12]. The fundamental autonomy of agents (their assumed group state, in which they acknowledge or even shield themselves from exterior influence) leads to the most basic form of alignment of roles, based on voluntary cooperation [12]. From this starting point, we can later understand how coercion and involuntary impositions attempt to simulate this configuration. The outcomes appear somewhat similar to exterior agents, but the uncertainties make be quite different.

Begin with the simplest directed relationship in figure 10. In this figure, one agent offers an attribute or service, by promising $+L$, and another avails itself of the service $-l$. We write this in conventional notation as:

$$A_L \xrightarrow{+L} A_S \quad (50)$$

$$A_S \xrightarrow{-L} A_L. \quad (51)$$

Implicit in this notation (and in Promise Theory) is the idea that there is no implication of force or obligation (necessity) in these expressions. Both agents are free to withdraw their promises at any time,

Figure 10: The simplest directed relationship which induces the beginnings of a hierarchy. One agent offers an attribute or service, by promising $+L$, and another avails itself of the service $-L$.

so no autonomy (causal independence) is violated. This is what distinguishes Promise Theory from modal logics and deontic formulations, which abound across the sciences, and form the dominant viewpoint on reason. By contrast, the more successful application of physics has continued to point to the importance of locality, or causal separation of influence in spacetime, which is expressed by the principle of autonomy in Promise Theory. However, locality introduces new conundrums, one of which is the mechanism by which localized agents express properties and effectively discriminate other agents of different types (from elementary charge at the subatomic level to job descriptions at a human level, etc). Promise Theory does not try to answer that, but rather invokes elementarily of promises as its semantic primitive.

The agents individually assess a kind of utility to the binding. Note, however, that the offer of something ‘intended’ (in this case, a service called L flows entirely in one direction). The invariance of L is what we mean by ‘intent’ in this case. Intent, in a promise theoretic sense, is a form of causal alignment.

Notice how, in the invited forms, e.g. (23), the conditionality of promises associated with the mandate is what implies a partial ordering of events in time, i.e. a notional direction which breaks the time symmetry and leads to authorization in time. Without that, there is semantic asymmetry and dynamical symmetry. Note also that a basic amount of ‘internal state’ or memory to remember past state can affect the assessment of an agent about whether a promise has been kept satisfactorily or not. Thus, we expect—as agents become more sophisticated on the scale from atoms to humans and ecosystems—that their judgements will likely also be more intricate and subtle. An agent which forgets past states (like a Markov process) always has the same level of trust in its counterpart, but an agent that holds a grudge for past digressions may be fickle. Thus, one expects primitive interactions to be somewhat invariant over long timescales, while complex agents with interior structure may have unpredictable and non-linear behaviours.

7 Summarial remarks

This paper proposes a definitive way of understanding authority as a universal phenomenon—composing the elementary interactions between autonomous agents and their semantics. Its aim is to unify many apparently different interpretations in common usage as a single application of Promise Theory. As a contribution to Social Science, it attempts to go beyond statistical or anecdotal evidence, to follow the Natural Science tradition based on structure of information—agent configuration and interaction. It doesn’t make particular assumptions about human nature, norms or social institutions, but neither does it attempt to explain the full range of dynamics from which authority waxes, wanes, evolves, or resonates with individuals. The complicated relationship with ‘power’ and related topics are deferred until future work.

Authority comes in many flavours. There is no single configuration of agents, nor interpretation of events that typifies the authority-subordinate relationship. Collective social phenomena may derive from networks of arbitrary complexity. What I have tried to show is that all such cases must have some necessary and sufficient features: authority is associated with a specialization of roles leading to a quasi-service relationship of some kind, such that a number of agents or ‘subordinates’ may begin to align with the service autonomous or ‘voluntarily’. The role of time order is not decisive, i.e. who came first, though timescale is key in interactions; nor is it necessarily about prior privilege, suitability, or competence. It’s not a race to win, rather it’s principally about the voluntary acceptance of roles with respect to the reference promise or service conferred by the authority, which is built in tandem with mutual trust. The reasons why agents accept such a breaking of symmetry between roles could be many and varied. There is plenty of scope to study these further in this framework.

What Promise Theory predicts is that a dynamics of authority lie in an individual perception of economic benefit—i.e. in the assessment of some ‘payoff’ in its most general meaning. Reader beware,

however, this statement does not necessarily imply a real valued von Neumann type utility metric [30] as the only way of selection. The depth and sophistication of assessments depends very much of the assessment capabilities of agents at various scales: a human has more scope for complex assessment than an amoeba or a machine, but the same underlying dynamic applies to all. Assessment may be seen through the lens of optimization, morality, group compliance, and so forth. In short:

- An authority (as an agent) is basically a trusted party.
- The emergence of its authority (as a promise) is an outcome of symmetry breaking between agents' characteristic promises, as a result of interior processes. One such promise is identity or title, e.g. family background, but in general a more functional role is the arbiter of distinguishable identity rather than an agent's proper name. Promise Theory predicts that some agents will not be able to distinguish between different possible authorities and therefore may appoint several or appear to choose irrationally.
- If there is no individual and direct mandate, then agents may abdicate from their subordinate positions and choose a different path, which is not aligned from a single source, e.g. 'go AWOL'.
- Regardless of whether an agent accepts a mandate or appointment as an authority, any agent is free to appoint its own authorities on different matters.
- Lying, deceptions, and threats are kinds of promise. They can apply both to the mandate promise and the appointed service (leadership) promise. In both cases, the promises also have to be accepted—even after the mandate has been given. So voting someone in is a necessary condition but not a sufficient one.
- When an agent authorizes some other's behaviour, one should be sceptical, as that is a promise on behalf of another—which violates the tenet of basic autonomy. Once formed, authority may persist on its own by force of habit or systemic inertia, as the underlying assessments and promises may be shrouded in layers of indirection and may thus take some time to assess and decay.

There is plenty of room for extending the method to address additional matters. In the spirit of the natural sciences, the goal here has been to find a suitably idealized approximation which can be used as tool for predictive reasoning. In closing, the reader could apply the derived concept of authority to this document. Is it authoritative on the subject of authority? Might one take a vote to secure a mandate, or merely assess the emergence of promises made individually? I'll leave that assessment to the reader, and defer other questions to future work.

Acknowledgement: I'm grateful to Jan Bergstra, Mark Haugaard, and Chuck Pezeshki, and Jim Rutt for generously providing detailed comments and criticisms over the past year.

References

- [1] J.A. Bergstra and M. Burgess. *Promise Theory: Principles and Applications (second edition)*. χ tAxis Press, 2014,2019.
- [2] D. Morselli and S. Passini. New perspectives on the study of the authority relationship: Integrating individual and societal level research. *Journal for the Theory of Social Behaviour*, 41(3):291–307, 2011.
- [3] M. Haugaard. What is authority? *Journal of Classical Sociology*, 18(2):104–132, 2018. DOI: 10.1177/1468795X17723737.
- [4] M. Weber. *Economy and Society: An Outline of Interpretive Sociology* (ed G. Roth and C. Wittich). University of California Press, 1978.
- [5] M. Haugaard. *The Four Dimensions of Power: Understanding Domination, Empowerment, and Democracy*. Manchester University Press, 2020.
- [6] S. Galam. *Sociophysics*. Springer, 2012.
- [7] J. Bergstra and M. Burgess. *Promise Theory Case Study on the 2016 Brexit Vote*. χ t-axis Press, 2017.

- [8] J. Bergstra. *Promises and Threats by Asymmetric Nuclear Weapon States*. χ t-axis Press, 2019.
- [9] J. Bergstra and M. Burgess. *Money, Ownership, and Agency*. χ t-axis Press, 2019.
- [10] J.A. Bergstra and M. Burgess. A promise theoretic account of the boeing 737 max mcas algorithm affair. *arXiv:2001.01543 [cs.OH]*, 2019.
- [11] J.A. Bergstra and M. Burgess. Candidate software process flaws for the boeing 737 max mcas algorithm and risks for a proposed upgrade. *arXiv:2001.05690 [cs.CY]*, 2019.
- [12] M. Burgess. Spacetimes with semantics (ii). *arXiv.org:1505.01716*, 2015.
- [13] J. Searle. *Intentionality*. Cambridge, 1983.
- [14] J. Searle. *Speech Acts*. Cambridge, 1969.
- [15] F. Flores. *Conversations for Action and Collected Essays*. (independently published), 2012.
- [16] J.A. Bergstra and M. Burgess. Local and global trust based on the concept of promises. Technical report, *arXiv.org/abs/0912.4637 [cs.MA]*, 2006.
- [17] J.F. Nash. *Essays on Game Theory*. Edward Elgar, Cheltenham, 1996.
- [18] R.B. Myerson. *Game theory: Analysis of Conflict*. (Harvard University Press, Cambridge, MA), 1991.
- [19] D. Mezick and M. Sheffield. *Inviting Leadership: Invitation-based Change*. Freestanding Press, 2018.
- [20] R. Axelrod. *The Complexity of Cooperation: Agent-based Models of Competition and Collaboration*. Princeton Studies in Complexity, Princeton, 1997.
- [21] R. Axelrod. *The Evolution of Co-operation*. Penguin Books, 1990 (1984).
- [22] M. Burgess. Spacetimes with semantics (i). *arXiv:1411.5563*, 2014.
- [23] M. Burgess. Spacetimes with semantics (iii). *arXiv:1608.02193*, 2016.
- [24] M. Burgess and S. Fagernes. Voluntary economic cooperation in policy based management. *IEEE Transactions on Network and Service Management*, page (submitted).
- [25] M. Burgess and S. Fagernes. Laws of human-computer behaviour and collective organization. *submitted to the IEEE Journal of Network and Service Management*, 2008.
- [26] R. Dunbar. *Grooming, Gossip and the Evolution of Language*. Faber and Faber, London, 1996.
- [27] W.X. Zhou, S. Sornette, R.A. Hill, and R.I.M. Dunbar. Discrete hierarchical organization of social group sizes. *Proc. Royal Soc.*, 272:439–444, 2004.
- [28] C.E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, 1949.
- [29] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. (J.Wiley & Sons., New York), 1991.
- [30] J.V. Neumann and O. Morgenstern. *Theory of games and economic behaviour*. Princeton University Press, Princeton, 1944.