An Operational Definition of Institutional Beliefs

Cuizhu Wang[1]
Jagiellonian University
cuizhu.wang@uj.edu.pl

Simon Graf[2]
University of Vienna
prsimongr@gmail.com

Konrad Werner[3]
University of Warsaw
k.werner2@uw.edu.pl

- Penultimate Draft: please cite published version -

Abstract:
Some of our beliefs are institutional; that is, beliefs whose content is to a large extent shaped by institutions, such as beliefs about intellectual property, trade policy, or traffic rules. In this chapter, we propose a novel account of institutional beliefs, as we call them. In particular, we argue that institutional beliefs are primarily attributable to social entities, such as groups or collectives, and only secondarily to individual agents. This is because institutional beliefs respond to specific problems that, in principle, only collectives face. However, existing methodology in behavioural science primarily relies on eliciting beliefs through individual choice data, which presents challenges for empirically modelling our proposed concept of institutional beliefs. To address this challenge, we propose to speak of institutional beliefs attributable to individual agents as 'qua beliefs'. That is, institutional beliefs attributable to individual agents acting *qua* members of certain institutionally shaped groups. Within this framework, individuals' qua actions occur in a shared realm of problems that can only be addressed through collective efforts. The relationship between individuals and their groups can be rendered tractable under formalised conditions. Under these conditions, choice data collected from individuals need not obstruct the elicitation of institutional beliefs for empirical study at both the collective and the individual level. We refer to these conditions as "IS conditions", which are developed based on the synthesis between an interpretationist understanding of mental states and standard models of revealed preference theory as used in experimental economics. For an illustration of how this approach can be operationalised experimentally, we direct our readers to chapter 10 of this volume, where an experimental study applied our account of institutional beliefs.
Keywords: *Institutional Beliefs; Interpretationism; Institutions; Divergence Arguments; Intentional Stance Conditions; Operational Definition*

0. Introduction

Some of our beliefs are shaped by the wider institutional structures in which we are embedded. For example, the expectation that our draft chapter is safe to circulate among colleagues is shaped by institutions that recognise and protect intellectual property. We believe that we can get valuable feedback from our colleagues without risking our draft being taken over by someone else, simply by putting their name to it. These kinds of institutionally shaped beliefs can be called institutional beliefs.

Moreover, recent work in social ontology suggests that not only individuals but also social entities, thus collectives or groups of various kinds should sometimes be understood as genuine believers, having

[1] Dr. Cuizhu Wang is an Assistant Professor at the Faculty of Law and Administration, Jagiellonian University. Her ORCID number is 0000-0002-4703-3323. Her contribution to this chapter was funded by the National Science Centre, Poland (grant no. 2020/37/B/HS5/00272).
[2] Dr. Simon Graf is currently a research fellow at the Institute Vienna Circle, University of Vienna. His ORCID number is 0000-0003-2235-8639.
[3] Dr habil. Konrad Werner is a Professor of philosophy at the Faculty of Philosophy, University of Warsaw, Poland and PTH College Brixen, Italy (visiting). His ORCID number is 0000-0002-8947-3722. His contribution to this chapter was funded by the National Science Centre, Poland (grant no. 2022/45/B/HS1/02822) and by Provincia autonoma di Bolzano – Alto Adige (no. 13585/2023).

collective-level beliefs. If those collective beliefs can be shaped by institutional structures as well, then not only individuals but also groups can have institutional beliefs. For example, if a trade union demands higher minimal wages and directs this call to state authorities, by doing so the union as a collective articulates a belief, shaped by the more than a hundred years old, institutions of the welfare state.

But what exactly are institutional beliefs? And, if those beliefs exist on the individual and the collective level, what is the relationship between the institutional beliefs we as members of collectives have and the institutional beliefs these collectives have themselves? Our chapter attempts to answer these questions with a special concern on providing an answer that is *operationalisable*, i.e., an answer that can be employed in experimental settings which test the existence of institutional beliefs. Accordingly, our interest is both theoretical, as well as empirical. Operationalisability is the key desideratum which we then use to navigate and discuss a wider scope of theoretical considerations concerning the nature of (collective) beliefs and institutions in general, as well as institutional beliefs in specific.

However, the condition of operationalisability faces a challenge because, as we will argue, institutional beliefs are primarily attributable to groups and only secondarily to individual agents. This is because institutional beliefs respond to specific problems that, in principle, only collectives confront, such as the problem of adopting a certain trade policy that is recognised by the members of a certain economic sector. In order to address this challenge, we propose to speak of institutional beliefs that can be attributed to individual agents qua members of a certain institutionally shaped group and qua acting in a shared domain of problems that can only be solved through collective efforts.

This framing situates individual-level behaviours, insofar as they manifest institutional beliefs, within a shared problem domain. Within this domain, collections of individuals can be treated as bearing institutional beliefs at the collective level. This then raises the question: under what conditions can such collections of individuals be regarded as collectively holding the institutional belief? This question leads us to define three conditions that formally establish operationsability. Drawing on an interpretationist view of belief and based on its synthesis with the revealed preference theory, these conditions enable the empirical elicitation of institutional beliefs.

Here is the outline. Section [§1] gives an overview of different understandings of collective beliefs and motivates the idea that some collectives, such as structured groups, can have genuine beliefs. However, we remain neutral on ontological questions about the nature of collective beliefs. Instead, we opt for an interpretationist strategy of belief attribution that satisfies some minimal operationalisation requirements. This will be important given the challenge outlined above. In [§2] we turn to the concept of an institution. In particular, we discuss the enabler account of institutions and its erotetic extension proposed by Werner (2023). In [§3] we bring together insights from the previous two sections to explain what we understand by an institutional belief. We then argue that these group-level institutional beliefs can be tested by studying individual choice and belief data. As just noted, this is a consequence of our interpretationist methodology, which is well aligned with the revealed preference theory used in experimental economics.

In [§4] we present the details of the proposed operationalisation. In particular, we argue that in order to count as having institutional beliefs, an agent's observable behaviour (for modelling) must satisfy the intentional stance conditions (IS conditions).

1. Collective Beliefs

Beliefs and other doxastic attitudes are frequently ascribed not only to individual agents but also to collectives, such as small-scale groups, communities, or other social entities.[4] We read that 'the college union thinks that the industrial action is going to plan' or that 'the IPCC believes that the currently observed climate change is predominantly anthropogenic'.[5] Among social philosophers, it is nowadays common to take at least some of those ascriptions to be non-metaphorical, genuine attributions of doxastic states. In other words, many social philosophers reject eliminativism, that is to say, the view that "non-metaphorical group attitude talk is always misconceived" (List & Pettit 2011: 7).

Broadly speaking there are two different *non-eliminativist* approaches to understanding the nature of collective attitudes (List & Pettit 2011; Lackey 2021: ch.2). So-called *summativists* think that ascribing an attitude to a collective is to indirectly ascribe it to its members. For example, to say that group G believes that *p* is to say that a sufficient portion of some relevant members believes that *p*. Others called *non-summativists* have argued that collectives can have doxastic attitudes that are independent or distinct from the doxastic attitudes of their members.

Non-summativism comes in two different forms: *deflationism* and *inflationism*. While inflationists think that collective-level attitudes can be independent from or 'over and above' individual-level attitudes, deflationists, on the other hand, think that there is some direct (but non-summative) relationship between happenings on the individual-level and the collective-level. So, in contrast to summativists, who make a very specific reductionist claim, for deflationary but non-summative views, the relationship between individual-level and collective-level attitudes can be more complex; so, for example, "group-level beliefs may be reducible to member-level behaviour more broadly, and therefore depend on other member-level attitudes such as member-level acceptance states" (Graf 2023: 241-242).[6]

In sum, we obtain the following structure of possible views on collective-level attitudes, depicted in the following tree diagram as shown in Figure 1:

[4] We decided to primarily use the term 'collective belief' rather than 'group belief' since it better serves as a neutral umbrella term. In particular, whenever we talk of 'collective belief' we mean to include so-called 'group beliefs' as well as beliefs held by other kinds of social entities. For further discussions of the difference between group beliefs and collective beliefs see, for example, Brouwer, Ferrario and Porrello (2021: §2.3).

[5] These examples are taken from Graf (2023).

[6] Note that, as pointed out by Habgood-Coote (2019: 923) and Graf (2023: 241-243), the terms deflationism and inflationism are often used interchangeably in the literature.

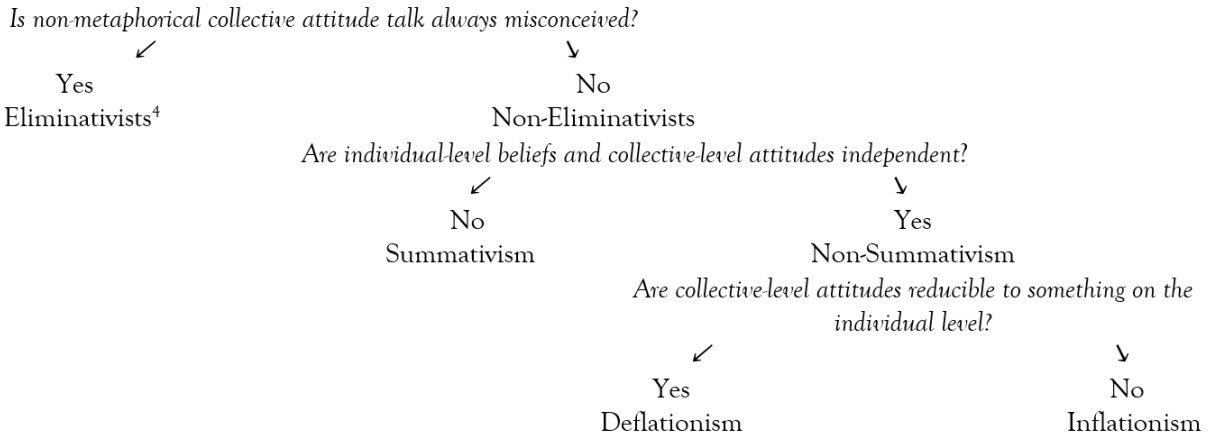

Figure 1

Having this classification of views at hand is helpful since it enables us to localise our proposal as to how institutional beliefs, which we take to be a specific kind of collective belief, can be operationalised for the sake of empirical research. After all, as we will argue below [§3], we think that institutional beliefs are first and foremost collective-level beliefs and only secondarily held by individuals. However, since our primary concern is operationalisability we need to have a conception of collective-level beliefs that allows us to systematically test those beliefs on individuals in controlled experimental settings. We can formulate this desideratum as follows:

> **Operationalisability Desideratum**: A theory T of collective-level attitudes is *operationalisable* iff according to T, to ascribe an attitude D to a collective G is to (i) interpret the behaviour of some of G's member's and/or (ii) make certain testable predictions about the behaviour of (some of) G's members.

Since this *Operationalisability Desideratum* is a *methodological* desideratum, this raises the question which of the above-listed *ontological* understandings of collective-level beliefs are operationalisable in the desired sense. A first attempt to answer this question is to point towards one prominent way to distinguish deflationary from inflationary theories of collective belief. It is often said that the former but not the latter subscribe to *methodological individualism*, the view that we "should not postulate any social forces other than those that derive from the agency of individuals" (List & Pettit 2011: 3). Accordingly, a natural suggestion would be to narrow down the set of possible theories to all theories that subscribe to the following commitment:

> **Minimal Deflationist Commitment:** A collective G has the attitude D towards p only if at least one member m of G has the doxastic attitude D or the corresponding pro-attitude D*.

Both summativist views (Quinton 1975), partly summative views (Lackey 2021; §2), as well as non-summative but deflationary views, subscribe to this commitment since the parenthetical 'pro-attitude' is a placeholder for anything on the individual/member level that is required for constituting the corresponding collective-level attitude *D*. This could be certain individual-level judgements as in premise-based judgement aggregation views (List & Pettit 2011), some *qua* group member attitudes (Kallestrup 2024), or other pro-relations such as the individual's commitment towards *p* required for the collective's joint commitment towards *p* (Gilbert 1987; Tuomela 2004). In other words, most extant accounts of collective attitudes, including summativist and non-summativist subscribe to this minimal deflationist commitment.[7] And since this commitment suffices for operationalisability, most extant accounts of collective attitudes are operationalisable in the desired sense.

However, while highlighting this minimal deflationist commitment enables us to demonstrate that our approach is in principle compatible with a wide range of extant views, we do not want to commit ourselves to any of the above-listed conceptions nor to the minimal deflationist commitment they have in common.[8] After all, our main concern is operationalisability. Accordingly, we leave these ontological questions aside and turn directly to the question of how we attribute beliefs to individuals and various kinds of collectives such as groups.

While the whole debate sketched above concerns the nature of collective beliefs, one might wonder how the very attribution of beliefs to anybody or anything, including groups and individuals, works. To address this puzzle, we adopt the interpretationist understanding of mental states championed by Dennett (1987, 1991).[9] Interpretationism can be understood as, roughly speaking, the view that to be a genuine believer is to be *reliably interpretable* from the *intentional stance*. If we take the intentional stance towards an entity, we attribute beliefs and desires to this entity and then predict how it will rationally behave given those beliefs and desires. As such, the intentional stance contrasts with other stances, such as the *physical stance* and the *design stance*. While the physical stance involves predicting the behaviour of an entity by appealing to its physical states and our knowledge of physical laws, the design stance involves predicting the behaviour of an entity by appealing to its design or (biological) function.

Interpretationist approaches to collective beliefs argue that since we can interpret or predict the behaviour of some collectives in much the same way that we can interpret or predict the behaviour of individuals, namely from the intentional stance, collectives too can have genuine beliefs (Tollefsen 2002,

[7] Since we do not wish to adjudicate between these different views of collective belief for now, we invite readers to modify the condition in a way that suits their theoretical preferences. For a similar notion of pro-attitudes see Silva (2019: 275).

[8] In fact, other ontologically less commital theories might also be operationalisable in the desired sense, as long as they have testable systematic consequences on the individual level.

[9] For other influential interpretationist understandings of belief see Lewis (1974) or Davidson (1984). For a contemporary discussion see Williams (2020).

2015; List & Pettit 2011; Brouwer, Ferrario & Porrello 2021).[10] This allows us to spell out a criterion for believer identification (Backes 2021: 10324):

> **Interpretationism (identification criterion):** A system qualifies as a genuine believer as long as (i) its behaviour is reliably interpretable (or predictable) from the intentional stance and (ii) applying the intentional strategy to the system yields new predictive power that we didn't antecedently have.

Crucially, the identification criterion requires that we are not only (i) able to predict the behaviour of a system for it to count as genuine believer but also (ii) that we gain additional predictive power in doing so. That is, for collectives to count as genuine believers we need to get more predictive power when taking the intentional stance towards them. This is true of human agents. Usually, attempting to predict how human agents will behave from the design or physical stance is hopeless. By contrast, when we apply the intentional stance to a human agent, their behaviour usually falls into recognisable patterns that enable us to predict their future behaviour at least to some extent.

Tollefsen uses the following example to illustrate the predictive power of applying the intentional stance to collectives (2002: 402). Suppose we want to explain the incidents of gunfire during a naval blockade. We read the official rules of engagement published to govern Navy operations. These rules embody the Navy's rational point of view. We can then explain why the Navy sometimes fires at other ships. Under certain conditions, specified in its rules of engagement, the Navy will fire at anyone it believes to have hostile intent.[11]

This example does two things. It demonstrates how interpretationism works, but in addition to that it demonstrates the practical benefits of a non-eliminativist understanding of collective attitudes. First, if we take the intentional stance towards the Navy, we attribute beliefs, such as the belief that unauthorised ships entered their sovereign territory, and desires, such as the desire to protect itself against hostile intentions, to the Navy itself. This gives us the power to explain and predict various actions of the Navy better than if we were to apply a different attitude to them. Second, if we were to apply the intentional stance only to specific individuals, such as military commanders at different stages of the decision-making process, we would lose sight of the bigger picture, in which there are specific procedures that limit individual decision-making capacity, resulting from larger strategies adopted by other social entities of which the Navy is a part, such as the state.[12] Differently put, while one might try to explain the behaviour of some individual by taking the intentional stance towards them, this strategy will likely not be as effective in terms of predictive power given all the factors just listed.

[10] As pointed out by Brown (2024), List and Pettit (2011) express sympathy for both functionalism and interpretationism. Note, further that Brouwer, Ferrario and Porrello (2021) opt for a Lewis-style interpretationism rather than the Dennett inspired version that we or Tollefson (2002, 2015) rely on.

[11] For a further discussion of this example see Backes (2021).

[12] See, for example, Gray (1999) for an overview.

The explanatory power of using interpretationist strategies to explain the behaviour of social entities is a powerful argument for the existence of collective-level beliefs. This naturally links the interpretationist approach to a wider class of influential arguments against summative understandings of collective beliefs, so-called divergence arguments. Divergence arguments purport to show that member-level and collective-level attitudes can diverge (Lackey 2021: 56-58; Graf 2023: 241).[13] That is, for example, that a collective G can believe that p while none of G's members m_1-m_n does. The main job of divergence arguments is to provide examples supporting the possibility of divergence of collective attitudes and individual/member attitudes, as in the above-used Navy example.[14]

Note that this means arguments from interpretationism and divergence arguments may go hand in hand in motivating a deflationary understanding of collective beliefs. On an interpretationist picture, we get divergence cases whenever (i) we are able to reliably predict what a group that's behaving rationally is likely going to do without (ii) being able to reliably predict what some of its members are likely going to do assuming they too behave rationally.[15] To apply this to the Navy example, it seems that we can predict the Navy's behaviour much more effectively if we take the intentional stance directly towards the collective rather than to the individual members of the Navy separately.

All that being said, we remain neutral with respect to the specific metaphysical lessons about the nature of collective attitudes, we might be tempted to derive from these observations. That is, our main concern is the operationalisability and testability of a specific kind of collective belief ascription; that is, ascriptions of what we call institutional beliefs. However, before we can move on to spell out the operationalisation criteria for these institutional beliefs in detail [§4], we need to turn to the concept of institutions first [§2], which will lead us to motivate the underlying assumption that institutional beliefs primarily exist on the collective level [§3].

2. Problem domains and institutions

We have made clear how collective-level institutional beliefs fit into a broader context of different theories of collective-level attitudes in general. Furthermore, we have spelt out interpretationist criteria for attributing such attitudes, whether they refer to individuals or groups. The only thing still to be explained is the "institutional" qualification. That is to say, what it means for some beliefs to be institutional; and especially why some collective-level beliefs are of the institutional type.

[13] Divergence arguments have been used to motivate and support non-summativist accounts of group belief (Gilbert 1987), group assertion (Lackey 2021: 158-163), collective justification (Schmitt 1994; Mathiesen 2011; Silva 2019), as well as different kinds of collective knowledge (Bird 2014; Habgood-Coote 2019). For an overview see Graf (2023: 241-244).

[14] Note that divergence arguments merely demonstrate that collective-level attitudes are not reducible solely to the respective member attitudes, but not that they are not reducible at all. Accordingly, divergence arguments have not only been used to support inflationary but also deflationary non-summative accounts within social epistemology (Graf 2023; Habgood Coote 2019).

[15] For further discussion see Backes (2021: 10322-10323).

To make this more precise, we must start by clarifying what we mean by institution. Yet, we do not take any ontological position on the *nature* of institutions. In other words, we do not investigate what kind of being institutions are, but instead focus on the underlying epistemic, in particular belief-forming, functions of institutions. As such, the ontological concern serves as a background here, not a proper puzzle to be addressed.

There are two main positions on the ontological question of what an institution is. According to the most widespread view, known as the rule-following approach, institutions are formal or informal norms that govern behaviour. They are, as Douglas North (1990) famously puts it, the "rules of the game" (North 1990: 3) in a society. For example, there are certain normative constraints on who can marry whom, which side of the road to drive on, who can benefit financially from a song or a book, and so on. A particular problem with this account is that it conceives of norms as abstract and top-down constraints, whereas their implementation is always local and heavily dependent on local, bottom-up conditions. There is therefore a tension between the top-down and bottom-up levels. Ha-Joon Chang (2011), for example, argues that some top-down formal institutions can actually stifle local economic growth if people on the ground simply don't see the need for a particular norm, or if the norm goes against their perceived interests. Strict intellectual property laws imposed on developing countries, he says, are a good example of the latter scenario (see also Werner 2023).

Thus, the rule-following view has a hard time articulating and doing justice to local demand for institutions as rules: how it is generated and how it is responded to. From this perspective, the alternative view available, the so-called equilibrium approach, does a better job (see Hodgson 2006; Greif & Kingston 2011 or Guala 2016 for a comparative analysis and overview).

The equilibrium approach does not focus on a top-down norm, but rather on bottom-up, spontaneous behavioural regularities. Indeed, based on certain results from game theory, in particular the concept of Nash equilibrium, it is argued that complex and dynamic social systems exhibit a tendency towards unplanned coordination of expectations between agents, which leads them to adopt a certain pattern of behaviour sooner or later (see e.g. Binmore 2010). A common example is car traffic.

Now, we would like to highlight the fact that both of the invoked conceptions, as observed in Werner (2023), assume that the agents *know what they want*, they have their goals, and therefore they need specific institutions. The notion of demand reappears here, leading to the crucial question: how does it come about that people select certain things, states of affairs, relationships, etc. in their environment and take them *as* potential goals to be achieved, goods or services to be accessed, opportunities to be exploited, etc.? Is it always self-evident that a particular fragment of the world should be *aimed at* or pursued in a particular way? Werner (2023) argues that it's not but that the very capacity of an individual to aim at something requires an act of problematisation.

Problematisation transforms the environment of an agent into a realm of salient entities as if calling for action (see Werner 2024). This means problem spaces or problem domains. Hence, the capacity for problematisation transforms the environment into a problem domain filled with potential goals, challenges, opportunities, tasks, etc. To take a simple biologically inspired example, given the differences in their respective cognitive apparatus, bats live in different problem domains from hawks, and, going further, the problem domain of a hawk obviously differs from that of a wolf, in spite of the somewhat trivial fact that all of these animals may live in seemingly the same environment, in close proximity to one another.

Meanwhile, the two standard conceptions of institutions assume that *a* problem domain in which the agents do various things, constrained by top-down norms or bottom-up patterns, is already there; it's *given* to the agents, as if someone or something had prepared a scene on which institutions can play their roles of "constrains and enablers", to use Hindriks and Guala's (2019) phrase.

Here, Werner (2023) proposes that the emergence of a specific problem domain not only makes room for an institution and creates local demand for it; it may also be a *function of* an institution. In other words, institutions not only answer our concerns; they also decide, among other factors, what concerns we have, thus what problems we face. To put it even better – they decide on what we are able to perceive *as* a problem.

Good examples come from social and feminist epistemology. Think of the concept of sexual abuse. As Fricker (2007) argues, even quite recently in many societies unwanted sexual advances were conceptualised as mere flirtation, even if it caused discomfort in women subjected to this kind of behaviour. If norms governing behaviour are identified with institutions, it can be said that there were institutions allowing for the said kind of behaviour. But – crucially – the normative conceptualisation of such unwanted sexual advances *as* innocent flirtation also contributed to the lack of "interpretative resources," as Fricker (2007) calls it, that would enable women in these situations to express their own feelings better. In our terminology, this means that certain behaviours were not parts of the relevant problem domain because the institutions that were in operation in the said societies did not enable people to problematise these behaviours, i.e., to single them out from a huge pool of phenomena *as* something to be addressed; to see them *as* morally and practically relevant problems.

Hence, generally speaking, institutions are not only disablers or enablers or specific pursuits of problem-solving; first and foremost; they are enablers or disablers of the very recognition of something *as* a problem. This means that institutions have a certain epistemic function, which needs to be emphasised for the sake of further analysis.

3. Institutional Beliefs

People have plenty of beliefs that seem to be shaped by the formal and informal institutions they live within. This includes individual beliefs about marriage, gender, social roles, the roles of citizens, leaders and governments, banks and churches, etc. To take a very specific example from economic history, it seems likely that had there been no Navigation Acts enacted in England in the 17th century, the body of individual beliefs held by the people we call "mercantilists" would not have emerged in the form we know it, nor would the opposing body of individual beliefs held by Adam Smith and those we now call "free trade economists". This is because the said legislation re-shaped *not* only the rules of conduct but also the very problem domain in which a variety of people, including merchants, future policymakers, but also intellectuals, among many others, would operate. We refer to these latter beliefs as *institutional beliefs*.

However, what does this "shaping" of beliefs by institutions actually stand for? In other words, what specific subset of propositional attitudes does the category of *institutional* beliefs single out? To begin with, we propose to reflect on a difference between cases when we are interested in what an individual *as such* believes, e.g. what our friend John believes about tariffs (which is quite topical an issue for anyone raised with the mercantilist institutional framework) and cases when we are interested in what John believes *as* a holder of a certain institutionally shaped belief system called mercantilism. So, we might be concerned with what John *qua individual person* thinks about tariffs or with what John *qua mercantilist* believes about tariffs.

Admittedly, the difference is subtle, nonetheless quite significant. For imagine that you are about to debate with John. To prepare for the debate you must learn about him, which includes the roots of his beliefs, sources of arguments, motivations, etc. Now, if you cannot specify the political and intellectual tradition to which John belongs, your preparations must involve the entire complexity of his personal belief system. And this may be a mesh of both collective and individual, sometimes very private and intricate factors. Therefore, unpacking in sufficient detail in response to what problems John as a person has this or that conviction is an unlikely task.

However, if you can attribute John to a specific tradition, and you take him, for example, *as* a representative of the mercantilist belief system, you can see quite clearly what are the problems that his beliefs address, and, therefore, you can try to make the point that your beliefs answer them better; moreover, you can predict what John's argumentative strategy will be.

This is a very simplistic illustration of what is singled out by the subclass of beliefs we are interested in; namely, those beliefs that can be attributed to specified problems and solutions in a certain clearly addressable problem domain. Now, as proposed in the previous section, the very specification of problem domains in large part depend on the existing institutional framework. Therefore, the subclass at stake here, can be referred to as *institutional* beliefs.

One might argue that focusing specifically on the institutional beliefs one holds is somewhat artificial, since in reality people's beliefs are always influenced by a mix of social, institutional, but also private, emotional, biographical and many other factors. Therefore, cutting off the institutional part from the rest is a kind of surgery performed on the "living body" of one's belief system. What's the benefit from that?

We agree that the category does indeed single out only one aspect of real-life belief systems. Nevertheless, we suggest here that although it carries less information about real individual belief systems, specifying the institutional beliefs held by an individual believer, for example, our friend John, enables other believers to link at least some segments of John's real belief system to a well-recognised nexus of problems, demands and solutions. And this makes John's behaviour as an institutional believer more predictable.

Given the latter remark, the distinction can be articulated in the light of the interpretationist strategy adopted above [§1]. Namely, we claim that interpreting an individual agent holding the belief that *p* without specifying the institutionally determined problem domain that *p* addresses, and the problem-solving strategy to which *p* belongs, makes it difficult and sometimes even impossible to make reliable predictions about an agent's behaviour. In contrast, interpreting the agent as an institutional believer renders the agent remotely predictable.

However, all the remarks made so far were about *individual* institutional beliefs, which may seem to stand in tension with the remarks articulated above, that is to say, the adopted non-eliminativist understanding of collective attitudes. This is not the case though. Here is why.

A more standard debate on collective beliefs starts from the concept of belief whose prototypical exemplifications involve individuals, and then asks if the beliefs thought of this way can also be attributed to groups and other collectives. We propose a different route. Namely, institutional beliefs in our account primarily show up at the collective level. In other words, institutional beliefs are prototypical beliefs that collectives can be attributed with, which is characterised by the direct reference to a collective-specific problem domain. Thus it has better predictive power if we start at the collective level and then ask under what conditions the institutional beliefs can *also* be attributed to individual believers.

This is not supposed to mean that institutional beliefs can exist without individual believers. As specified in [§1], we do not take a stance on that. However, if we take a look at the cluster of theories which subscribe to the above-defined Minimal Deflationist Commitment, we can see that in each case of institutional belief *D* attributable to a collective G, there should be at least one member of G who holds *D* (or some respective pro-attitude *D** which indicates that G has attitude *D*). Otherwise, they could hardly be operationalised and empirically studied.

To wrap up this section, we propose that institutions have the function of binding collective-level beliefs (directly) as well as individual beliefs (indirectly, through the collective-level ones) to specific problem domains, so that they don't float around in the air, so to speak, and instead remain stick to the

particular goal or task. When interpreted as institutional, this whole dynamic of beliefs is tied to specific objects, states of affairs, services, etc., that the given community recognises as problems to be addressed.

4. The Intentional Stance Conditions for Operationalisation

The conceptual groundwork of the previous sections [§1] – [§3] provide us with a solid foundation for the ultimate goal of our investigations – an operationalisable understanding of institutional beliefs. Assuming interpretationism [§1] enables us to interpret groups and other collective entities as independent epistemic agents, facing their own institutionally-shaped problems [§2] and having their own doxastic attitudes directed towards these problems [§3]. We also showed how the prototypical collective-level institutional beliefs can connect with individual-level institutional beliefs. This suggests that the measurement of institutional beliefs consists primarily of the task of measuring them at the collective level, and secondarily at the individual level. This section addresses the central aim of this chapter: making our account operationalisable. We begin by revisiting the intentional stance and asking, 'How can we attribute beliefs to agents following the intentional stance?' Drawing on suggestions from Dennett's work and from other scholars who follow his approach, we define these insights as 'IS conditions'. Following this narrative elaboration, we introduce formal axioms developed by Savage (1972) with respect to the Revealed Preference Theory (RPT), which serves as a methodological application of the intentional stance. We conclude with a discussion on how synthesising between the IS and RPT supports the operationalisability of our account.

4.1 Intentional Stance Conditions

In [§1] we advocated an interpretationist understanding of mental states in support our account of collective-level beliefs in general and institutional beliefs in particular. However, how does the *intentional stance* actually work in reaching its two goals, interpretation and prediction? Dennett (1987: 17) summarises the procedure as involving the following steps:

> 1) first you decide to treat the object whose behaviour is to be predicted as a rational agent; 2) then you figure out what beliefs that agent ought to have, given its place in the world and its purpose. 3) Then you figure out what desires it ought to have, on the same considerations, and 4) finally you predict that this rational agent will act to further its goals in the light of its beliefs. A little practical reasoning from the chosen set of beliefs and desires will in many – but not all – instances yield a decision about what the agent ought to do; that is what you predict the agent *will* do.

Following Dennett's approach, we identify a set of conditions sufficient to "adopt" the intentional stance (IS). These four steps share three assumptions, which we summarise as three conditions in our work. Below, we briefly explain each of these IS conditions.

Condition 1 – Goal-orientated Rationality

First, we examine the underlying treatment of intentional systems as rational agents, leading to our first success condition (*C1*), termed 'goal-oriented rationality'. Dennett (1987: 49-50) describes the kind of rationality involved in applying the IS as follows:

> A system's beliefs are those it ought to have, given its perceptual capacities, its epistemic needs, and its biography. Thus in general, its beliefs are both true and relevant to its life. . . . (2) A system's *desires* are those it ought to have, given its biological needs and the most practicable means of satisfying them. Thus naturally evolved intentional systems desire survival and procreation, and hence desire food, security, health, sex, wealth, power, influence, and so forth, and also whatever local arrangements tend (in their eyes - given their beliefs) to further these ends in appropriate measure. . . . (3) A system's behaviour will consist of those acts that *it would be* rational for an agent with those beliefs and desires to perform.

What kind of rationality is this? In answering this question, there are several aspects to point out. First, it is goal oriented.[16] Dennett (1987: 50) argues that "we treat each other as if we were rational agents, and this myth – for surely we are not all that rational – works very well because we are pretty rational. This single assumption, in combination with home truths about our needs, capacities and typical circumstances, generates both an intentional interpretation of us as believers and desirers and actual predictions of behaviour in great profusion." In simple words, rationality as goal-oriented implies that rationality is not simply about being coherent, following logical rules, but rather about achieving one's goals and objectives in an effective way. This may involve making decisions that are not necessarily the most logical, but that are more likely to lead to the desired outcome (s).

This naturally links us to the adopted understanding of institutions and institutional beliefs. For institutions are – among other factors, of course – responsible for what goals individuals have; that is to say – in what problem domain they operate. And the category of rationality in the Dennettian sense makes sense only with respect to specified problems or goals, thus within a fixed problem domain.

Second, this view of rationality emphasises the importance of considering the context and circumstances in which decisions are made.[17] Dennett (1987: 20) suggests that "one starts with the ideal of perfect rationality and revises downward as circumstances dictate". Since circumstances change, this naturally links us to the postulate that beliefs, including institutional beliefs, are in fact best represented by problem-determined equivalence classes. The somewhat loose talk of "circumstances" gets more precise then. Such consideration of rationality requires a holistic attribution, which suggests seeing the choice context, the subject of all intentional attribution, and consideration of other beliefs and desires of the subject holistically (see Dennett 1987: 58). In short, according to Dennett, *any* behaviour an agent would

[16] Dennett uses 'goals' and 'desires' interchangeably, in this chapter we take the same stance.
[17] We elaborate it more in Condition 2.

perform given the beliefs and desires she has would count as rational. There is no other explanation of what counts as rational.

Condition 2 – Normativity

The second condition underlying a successful application of the IS is termed 'normativity' (C2). This condition requires that the IS associates the system's behaviours with its specific environment – biological, social, or cultural – under the assumption of rationality. Dennett (1987: 49) describes this as the system having what it '*ought to have,*' meaning the beliefs and desires it would ideally possess if perfectly adapted to its environmental niche.

This leaves us with a holistic understanding of normativity which we may, following Zawidzki (2007: 48), characterise as follows:

> (1) Intentional systems have the beliefs they **ought to** have, that is, true and relevant beliefs, given their perceptual capacities and informational needs;
> (2) Intentional systems have the desires they **ought to** have, given their biological needs;
> (3) Intentional systems behave in ways that count as rational **given** these beliefs and desires.

The normativity ('*ought to*') implies that to attribute propositional attitudes according to the IS is not to conjecture private mental states that are causally responsible for behaviour. Rather, it is to *locate behaviour in a normative frame* to allow the behaviour to count as reasonable given the agent's goals and available information (Zawidzki 2012, 2013). It also implies that the kind of beliefs ascribed due to the IS are not fiction or some hidden complex within an agent's head, but they are socially constructed.[18]

The consideration of circumstances in IS implies that the condition of normativity must be empirically measured. Dennett (1987: 49) suggests that in cases where a system is attributed with false beliefs, "special stories must be told to explain how the error resulted from the presence of features in the environment that are deceptive relative to the perceptual capacities of the system." Harrison and Ross (2023) reflect that the measurement of 'circumstances' in the Dennettian sense in economics is practised by relying on empirical evidence and particularly on experimental manipulation.

Let's take the currently most prominent account of social norms due to Bicchieri (2005, 2016) as an example. Bicchieri (2005, 2016) argues that individuals' behavioural responses are conditioned by networks of beliefs and preferences.[19] The theory suggests modelling norm-governed behaviour via measurement of socially engaged individuals' 1) empirical and normative beliefs about what others (in their community) will and should do, 2) the second-order empirical and normative beliefs about those

[18] Similarly, Harris and Ross (2023, 23-24) interpret the IS as a "product of cultural evolution" and regard Dennet's analysis of propositional attitudes as a scientific revision to "folk psychology".

[19] The account of preference that aligns with Bicchieri's model of social norms is revealed preference theory, which also aligns with the Dennettian intentional stance (see Wang 2022, 2023).

beliefs of others (in the community).[20] In addition, the model requires measuring individuals' reference network alignment. Wang (2022, 2023) argues that we must take the IS in making sense of the conceptualization of Bicchieri's theory. Once this is appreciated, we can see that Bicchieri's model aligns with the IS framework,[21] in the sense that it not only applies the IS attribution to observable behaviours, but also the requirement of reference network[22] alignment mimics the concept of 'circumstances' addressed by Dennett. In implementing Bicchieri's model, Wang (2022) designed a Normative Value Survey[23] to control for a normative reference network across the subject sample, in addition to adopting a demographic questionnaire. This cashes out as one instance of experimental manipulation of the condition of 'normativity', in comparison to *C2* in our account.

Condition 3 – Error Tolerance

We now turn to the third condition, 'error tolerance' (*C3*). This condition assumes that IS attribution requires a degree of error tolerance. Specifically, Dennett posits that the IS should yield predictions unattainable by other methods. This approach enables us to identify "the patterns in human behaviour that are describable from the IS, and only from that stance, and that support generalizations and predictions" (Dennett 1987: 25).

In order to make clear the reality of the patterns, Dennett (1991: 31) invites us to consider six frames (Figure 2)[24] as six subsets of data with different degrees of *signal to noise ratio*.

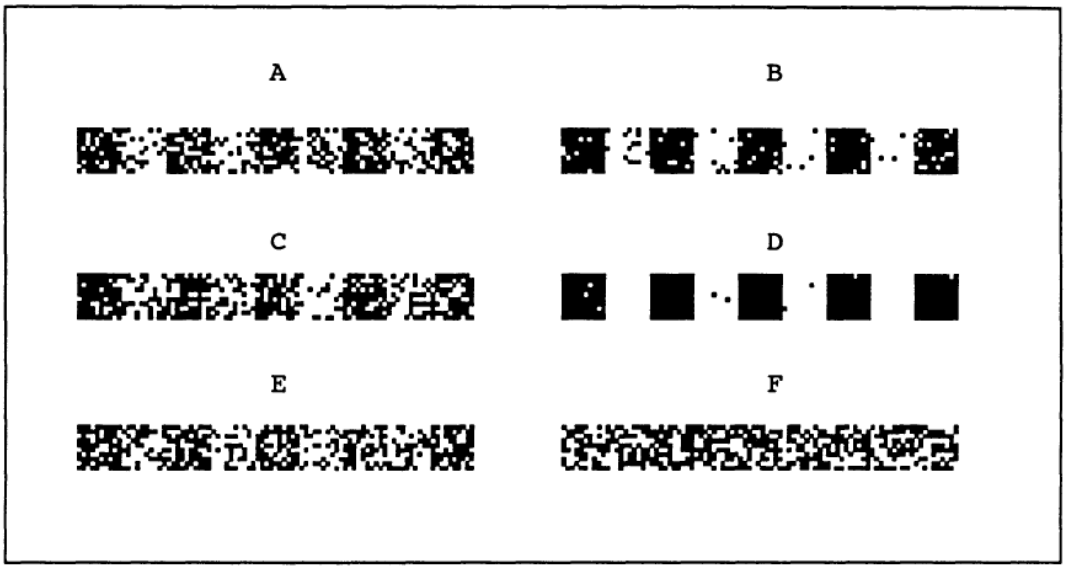

Figure 2

[20] The modelling requirement of these different kinds of belief according to Bicchieri refers to a concept of "belief consistency". Wang (2024, forthcoming) interprets this requirement as one of three conditions in operationalising Bicchieri's theory for norm modelling. See also Bicchieri (2006, 2017) and Wang (2022, chapter 2).
[21] See Wang (2022, 2023).
[22] An operational definition of the concept of reference network given by Ross and Wang (2024, forthcoming) refers to "the set of people whose normative expectations matter to the behavior and normative beliefs of an individual agent *i*."
[23] This survey is a selection of questionnaires from Afrobarometer questionnaires.
[24] Image from Dennett (1991: 32).

In the case where the pattern (see frame F in Figure 2) is indiscernible, it is simply that the noise ratio is beyond the limit for the pattern to be generated or recognised by human eyes. When we follow Dennett's example and consider our mind as the program generating patterns, it implies that the application of the IS requires an allowance for some degree of random error in interpreting and predicting the behaviour of a system. The success of such a practice depends on the noise ratio different brains can tolerate in pattern recognition.[25]

Again, there is a counterpart to this concept in contemporary experimental economics. In economics, random error is defined as 'stochastic error'. The primary importance of the stochastic structure of behavioural data is addressed in detail by Hey (2005), based on the pervasive feature of noise in choices which is well documented in economic literature (Schmidt & Hey 2004, Duffy & Smith 2023). The existence of noise not only requires statistical tools to be developed in accommodating analysis of error, but also it determines "the appropriate process of statistical inference to be carried out" (Hey 2005: 325). Harris and Ross (2016, 2023) developed structural models which provide stochastic specifications that attend to sampling errors in econometrics.[26]

Operationalising our account of institutional belief requires behavioural data to be analysed at both the collective level and the individual level. Therefore, the possible existence of behavioural noise in both procedures must be considered. Such need is addressed by Hey (2005: 327329), which therefore provides experimental leverages to our account of institutional beliefs.

In summary, *C1*, *C2*, and *C3* together are what Dennett specifies for adopting the IS. We argue that these conditions must be controlled for *altogether* when operationalising our account through empirical studies. For a detailed illustration, readers are encouraged to consult chapter 10 of this volume, which demonstrates how these three conditions are managed using experimental designs and statistical modelling to measure institutional beliefs at the collective level.

Specifically, separating vignette scenarios from the belief elicitation task, along with employing incentivised methodology, addresses Condition 1. Condition 2 is managed by a joint application of two surveys: the demographic questionnaire, and normative value survey,[27] with selected items from these instruments are transformed as covariates in estimating the marginal effects of the institutional beliefs. Finally, the error tolerance for Condition 3 is maintained at approximately 10%.

[25] The capacity for a specific perceiver to discern patterns depends on the subtle differences of the sense organs different creatures have. This is explained by the concept of pattern independence in Dennett (1991) and Ladyman and Ross (2007). In terms of the goal of this paper, we relax from the argument of this aspect.

[26] For example, Harrison and Ng (2016) practiced this understanding of the holistic nature of IS attribution to study behavioural Finance.

[27] This measurement is applied in the experimental study in chapter 10 as a shared with the measurement of social norms, following one of the conditions in Bicchieri's framework of social norms, interpreted by Wang (2022, forthcoming).

4.2 Synthesis between the Revealed Preference Theory and IS Conditions

The applicability of the IS condition in operationalising our account of institutional belief relies on the synthesis between the Revealed Preference Theory (RPT) and the IS (Ross 2005, 2014; Wang 2023). In economics, RPT suggests that we can ascribe a consistent set of *preferences* to an agent based on her behaviour. In representing preferences, utility stands for some numeric values agents assign to different possible options they face given their complex aims. Individuals' preference relations are often described through a utility function. The construction of utility functions is a mathematical device which allows for preferences to be modelled via choice data.

Various axiomatizations of RPT (Chambers & Echenique 2016) allow it to be utilised in empirical work (Harrison & Ross 2023). The first version, Weak Axiom of Revealed Preference (WARP) was due to Samuelson (1948) who showed the necessary mathematical conditions which enable the ascription of preferences to agents based on their behaviours without taking a position about underlying psychology. However, WARP wasn't useful empirically because it doesn't select a unique preference function from choice data. The second version, Strong Axiom of Revealed Preference (SARP) due to Houthakker (1950) selects a unique preference function from a specification of choices over all possible combinations of bundles of the objects of choice. However, the preference function according to SARP is defined over infinite sets of choices which renders SARP not very useful empirically. The third version, Generalised Axiom of Revealed Preference (GARP) due to Afriat (1967) guaranteed to find at least one preference function consistent with finite observed choices. GARP is the version which is mostly utilised in empirical work among economists because GARP guarantees to find at least one preference function consistent with finite observed choices. Following GARP, preferences are treated as summaries of actual choices (Harrison & Ross 2023).

As it follows, this section will introduce the axioms required to identify the utility functions that align with RPT. Then it will briefly discuss the synthesis between the IS and RPT.

4.2.1 Revealed Preference Theory (RPT)

RPT assumes that agents act consistently, and their behaviours are stable over time. Acting consistently involves agents taking risks at some times. Von Neumann and Morgenstern's (VNM) approach to utility, which is an additional restriction imposed on GARP, gives a primary role to risk. VNM utility measures how much an agent wants something by the relative magnitude of the risk she is willing to take to get it. Lottery choices are the common method for measuring subjects' risk preferences.[28] A range of formal theories are developed to identify processes that generate observable choice data in consideration of risk. The three core models of decision-making under *objective* risk are summarised by Harrison and Ross

[28] As pointed out by Harrison & Ross (2016) that lotteries over monetary rewards are typical ways but not the only ways experimental economists adopt in eliciting patterns in data.

(2016).[29] For our purpose, we demonstrate the axiomatisation of Expected Utility Theory (EUT) in demonstrating it as an application of the RPT.

EUT requires that the preferences over lotteries that an agent reveals in risky situations be consistent. Savage (1972) used EUT to develop the standard model of subjective probability, as an extension of RPT. According to EUT, an agent acts *as though* she believes that each possible state of the world in a decision problem has a *subjective* probability. Under such an assumption, agents' utility can be modelled *as if* there exists a subjective probability distribution *and* a utility function such that observed choices can be characterised as maximizing her Subjective Expected Utility (SEU).

The existence proof for identification of SEU requires satisfaction of the following axioms:

P.1 (Ordering) The relation ≤ is a simple ordering among acts.
If *F* is a finite set of acts, there exist *f* and *h* in *F* such that for all g in *F*.

$$f \leq g \leq h$$

This axiom defines the preference relation as a complete, transitive, and reflexive binary relation on *F*.

For the next axiom, where ~*B* is the complement of event *B*,

P.2 (Sure-Thing Principle) If acts *f*, g, and modified acts f', g' are such that:
1. in ~*B*, *f* agrees with g, and f' agrees with g',
2. in *B*, *f* agrees with f', and g agrees with g',
3. $f \leq g$;

then $f' \leq g'$.

This axiom is also called Sure-Thing Principle. Savage states that "if the person would not prefer *f* to g, either knowing that the **event** *B* obtained, or knowing that ~*B* obtained, then he does not prefer *f* to g. Moreover (provided he does not regard *B* as virtually impossible) if he would definitely prefer g to f, knowing that *B* obtained, and, if he would not prefer g to f, knowing that *B* did not obtain, then he definitely prefers g to f."[30] The Sure-Thing Principle implies that preferences are separable across events, hence it maintains an interpretation of the dynamic consistency of actions.

P.3 (Monotonicity) If $f = g$, and $f' = g'$, and *B* is not null; then $f \leq f'$ given *B*, if and only if $g \leq g'$.

This axiom states that the preference between two acts is state-independent, and that the preference between two acts depends solely on the consequences in states in which the payoffs of the two acts being compared are distinct.

P.4 (Weak Comparative Probability) If *f*, f', g, g'; *A*, *B*; f_A, f_B, g_A, g_B are such that:

1. $f' < f$, $g' < g$;
2a. $f_A(s) = f$, $g_A(s) = g$ for $s \in A$,
$f_A(s) = f'$, $g_A(s) = g'$ for $s \in$ ~A;
2b. $f_B(s) = f$, $g_B(s) = g$ for $s \in B$,
$f_B(s) = f'$, $g_B(s) = g'$ for $s \in$ ~*B*
3. $f_A \leq f_B$;

then $g_A \leq g_B$.

[29] The three formal theories referred here are Expected Utility Theory (EUT), Rank-Dependent Utility (RDU), and Cumulative Prospect Theory (CPT).
[30] Leonard J. Savage, *The Foundations of Statistics* (New York: Dover Publications, Second Edition, 1972), 21-22.

Axiom four requires that betting preferences be independent of the specific consequences that define the bets.

Three further axioms required by Savage, Nondegeneracy, Small Event Continuity, and Uniform Monotonicity, are needed only for technical reasons, and need not be specified for interpretations of later work in this article.[31]

4.2.2 The Synthesis between the IS and the RPT

RPT synthesises with the IS not only as a philosophical account, but also as a methodological implication of the IS (Ross 2014, Harrison & Ross 2023, Wang 2023). For our purpose, this section briefly explains the latter.

First, RPT and the IS require a similarly thin assumption of rationality (*C1*). To the IS, any behaviour displayed by an agent who holds the IS-fashioned beliefs and desires would count as rational. Whereas RPT assumes that agents act consistently, and their behaviours are stable over time. The assumption of rationality in both models take as given that there is a systematic relationship between agents' choices and their goals – a real pattern as Dennett (1991) calls it.

Second, RPT also synthesises with the IS condition of normativity (*C2*) in the following way. The condition of "normativity" suggests locating agent's behaviour in an environment where the behaviour can count as reasonable. This poses the need for assessing her "circumstances" in modelling her behaviour applying the IS. The 'small world' assumption of Subjective Expected Utility,[32] as an extension of the RPT synthesises this assumption in the sense that both models recognise the normative constraints which limit agents' goals and determine what agents believe and desire given those goals. In this world, it is arguably possible for an agent to take into account in advance of the implications of all conceivable future information for her subjective utility. Ross (2014: 505) interprets the 'small worlds' as worlds in which "institutionalized constraints tightly limit agents' goals and narrow the domains of the beliefs and conjectures that matter to their actions."

Third, both theories (i.e., the IS and RPT) do not rule out error (*C3*) in the modelling process. It requires only stochastic dominance of choices that produce "better" outcomes in expectation (Schmidt & Hey 2004, Hey 2005, Duffy & Smith 2023).

The synthesis between IS and the RPT shows that the two theories do not only share some common philosophical position, but also, we can understand the RPT as a methodological application of the IS. Therefore, RPT provides the operational leverage of any account of beliefs that relies on IS, such as our account of institutional belief.

---

[31] For details, see Savage (1972: 18-31).

[32] Subjective Expected Utility theory is one of the formal models of decision-making under *objective* risk, applying the RPT. The other two models Rank-Dependent Utility (RDU), and Cumulative Prospect Theory (CPT), See Harrison & Ross (2016).

In sum, based on the definition of institutional belief developed in this chapter, we argue that it can only be discerned from the point of view of one who adopts a certain predictive strategy, and its existence can be confirmed only by an assessment of the success of that strategy. The strategy we believe should be adopted is the Intentional Stance (IS) developed and championed by Dennett (1987). For operational purposes, we argue the application of the strategy aiming at empirical modelling of institutional belief must meet the IS conditions. The IS conditions suffice the goal of our account on two grounds: i) it aligns with the conditions conceptually developed in [§3], and ii) it aligns with the methodological framework of Revealed Preference theory (RPT) in allowing for the concept to be modelled via observed choice data.

5. Conclusion

In this chapter we have proposed a new understanding of institutional beliefs that can be operationalised in experimental settings, drawing on tools from experimental economics. In particular, we followed in the footsteps of those who have argued that it is possible to merge contemporary Revealed Preference Theory with the interpretationist understanding of mental states. Furthermore, we argued that the best way to rationalise the behaviour of individuals in institutional settings is to see them as acting qua members of a collective that holds a particular institutional belief. Accordingly, we suggest that while institutional beliefs are primarily held by social entities, they have testable consequences at the level of members' behaviour. Nevertheless, when this behaviour is collected from individuals who are members of the collective that satisfy the IS conditions specified in this chapter, it is possible to 'go back' to the collective level, so to speak, because these beliefs are shaped by problem domains that can only be faced by collective entities as a whole. For a concrete demonstration of how this account of institutional beliefs can be operationalised primarily at the collective level, we refer readers to chapter 10 of this volume, where an experimental study exemplifies our proposed framework.

6. References

Backes Marvin, 'Can Groups be Genuine Believers? The Argument from Interpretationism' (2021) 199 *Synthese* 10311.

Bicchieri Cristina, *The grammar of society: The nature and dynamics of social norms* (CUP 2005).

Bicchieri Cristina, *Norms in the wild: How to diagnose, measure, and change social norms* (OUP 2016).

Binmore Ken, 'Game theory and institutions' (2010) 38 Journal of Comparative Economics 245.

Bird Alexander, '"When Is There a Group that Knows?": Distributed Cognition, Scientific Knowledge, and the Social Epistemic Subject' in Lackey Jennifer (eds.), *Essays in Collective Epistemology* (OUP 2014).

Brouwer Thomas and Ferrario Roberta and Porello Daniele, 'Hybrid collective intentionality' (2021) 199 *Synthese* 3367.

Brown Jessica, *Groups as Epistemic and Moral Agents* (OUP 2024).

Chang Ha-Joon, 'Institutions and economic development: Theory, policy and history' (2011) 7 *Journal of Institutional Economics* 473.

Davidson Donald, *Inquiries Into Truth and Interpretation* (OUP 1984).

Dennett Daniel, *The Intentional Stance* (*MIT Press* 1987).

Dennett Daniel, 'Real Patterns' (1991) 88 *Journal of Philosophy* 27.

Duffy Sean and Smith John, 'Stochastic Choice and Imperfect Judgments of Line Lengths: What Is Hiding in the Noise?' (2023) *SSRN 4362781*: https://dx.doi.org/10.2139/ssrn.4362781.

Gilbert Margaret, 'Modelling Collective Belief' (1987) 73 *Synthese* 185.

Graf Simon, 'Permissive Divergence' (2023) 53 *Canadian Journal of Philosophy* 240.

Gray Colin S, *Modern Strategy* (OUP 1999).

Greif Avner and Kingston Christopher, 'Institutions: Rules of equilibria' In N. Schofield, & G. Caballero (eds.), *Political Economy of Institutions, Democracy and Voting* (Springer 2011).

Guala, Francesco 'The logic of normative falsification: rationality and experiments in decision theory' (2000) 7 *Journal of Economic Methodology*, 59.

Guala Francesco, 'The Normativity of Lewis Conventions' (2013) 190 *Synthese* 3107.

Guala Francesco, 'The Normativity of Institutions' (2015) 9 *Phenomenology and Mind*, 118.

Guala Francesco, *Understanding institutions* (PUP 2016).

Habgood-Coote, Joshua 'Group Knowledge, Questions, and the Division of Epistemic Labour' (2019) 6 *Ergo: An Open Access Journal of Philosophy* 925.

Harrison Glenn W and Ng Jia Min, 'Evaluating the expected welfare gain from insurance' (2016) 83 *Journal of Risk and Insurance* 91.

Harrison Glenn W and Ross Don, 'The psychology of human risk preferences and vulnerability to scare-mongers: experimental economic tools for hypothesis formulation and testing. (2016) 16 *Journal of Cognition and Culture* 383.

Harrison Glenn W and Ross Don, 'Behavioral welfare economics and the quantitative intentional stance' (2023) In Harrison Glenn W and Ross Don *Models of risk preferences: Descriptive and normative challenges* (Vol. 22) (Emerald Publishing Limited 2023).

Hey John 'Why We Should Not Be Silent About Noise' (2005) 8 *Experimental Economics* 325.

Hindriks Frank and Guala Francesco, 'The functions of institutions: etiology and teleology' (2019) 198 *Synthese* 2027.

Hodgson Geoffrey M 'What are institutions?' (2006) 40 *Journal of Economic Issues* 1.

Kallestrup Jesper 'Nonreductive Group Knowledge Revisited' (2024) 21 *Episteme* 565.

Lackey Jennifer, *The Epistemology of Groups* (OUP 2021).

Ladyman James and Ross Don, *Everything Must Go: Metaphysics Naturalized* (OUP 2007).

Lewis David, 'Radical interpretation' (1974) 23 *Synthese* 331.

List Christian and Pettit Philipp, *Group Agency: The Possibility, Design, and Status of Corporate Agents* (OUP 2011).

Mathiesen, Kay 'Can Groups Be Epistemic Agents?' In Schmid H. B., Sirtes D., and Weber M. (Eds.), *Collective Epistemology* (Ontos Verlag 2011).

North Douglass, *Institutions, institutional change and economic performance* (CUP 1990).

Quinton Anthony, 'Social objects' (1976) 76 *Proceedings of the Aristotelian Society*, 1.

Ross Don, *Economic theory and cognitive science: Microexplanation*. (MIT Press 2005).

Ross Don, *Philosophy of Economics* (PPT 2014).

Savage Leonard J, *The Foundations of Statistics* (Dover Publications1972 [1954]).

Schmidt Ulrich and Hey John D., 'Are preference reversals errors? An experimental investigation' (2004) 29 *Journal of Risk and Uncertainty* 207.

Schmitt Frederik, 'The Justification of Group Beliefs' in Schmitt Frederik (Eds.), *Socializing Epistemology: The Social Dimensions of Knowledge* (Rowman & Littlefield 1994).

Silva Paul, 'Justified Group Belief is Evidentially Responsible Group Belief' (2019) 16 Episteme 262. https://doi.org/10.1017/epi.2018.5.

Smith Vernon L, *Rationality in Economics: Constructivist and Ecological Forms* (CUP 2007).

Tollefsen Deborah, 'Organizations as true believers' (2002) 33 *Journal of Social Philosophy* 395.

Tollefsen Deborah, *Groups as agents* (Polity Press 2015).

Tuomela Raimo, 'Group Knowledge Analyzed' (2004) 1 *Episteme* 109.

Wang Cuizhu 'Normative Expectations and Subjective Beliefs: An Incentivised Experimental Study' (2022) PhD thesis. University College Cork. https://hdl.handle.net/10468/14502.

Wang Cuizhu 'Belief distributions and the measure of social norms' in Viciana, Hugo, Gaitán, Antonio, & Aguiar Fernando (Eds.), *Experiments in Moral and Political Philosophy* (Routledge 2023).

Werner Konrad, 'Who Asks Questions and Who Benefits from Answers: Understanding Institutions in Terms of Social Epistemic Dependencies' (2023) *Erkenntnis*.

Werner Konrad, 'Confined Truths and Cognitive Ecologies: When the Social Pursuit of Questioning Becomes Unreliable' (2024) *Topoi*.

Williams J Robert G, *The Metaphysics of Representation* (OUP 2020).

Zawidzki Tadeusz W, *Dennett* (Oneworld Publications 2007).

Zawidzki Tadeusz W, 'Unlikely allies: embodied social cognition and the intentional stance' (2012) 11 *Phenomenology and the Cognitive Sciences* 487.

Zawidzki Tadeusz W, *Mindshaping: A new framework for understanding human social cognition* (MIT Press 2013).