Synthese
https://doi.org/10.1007/s11229-021-03083-x

Salience reasoning in coordination games

Julius Schönherr[1]

Received: 11 February 2020 / Accepted: 11 February 2021

Abstract

Salience reasoning, many have argued, can help solve coordination problems, but only if such reasoning is supplemented by higher-order predictions, e.g. beliefs about what others believe yet others will choose. In this paper, I will argue that this line of reasoning is self-undermining. Higher-order behavioral predictions defeat salience-based behavioral predictions. To anchor my argument in the philosophical literature, I will develop it in response and opposition to the popular Lewisian model of salience reasoning in coordination games. This model imports the problematic higher-order beliefs by way of a 'symmetric reasoning' constraint. In the second part of this paper, I will argue that a player may employ salience reasoning only if she *suspends* judgment about what others believe yet others will do.

Keywords Coordination · Salience · Game theory · Belief suspension · Common knowledge

1 Introduction

There is a long line of philosophical research, inaugurated perhaps by the publication of Thomas Schelling's *The Strategy of Conflict* (1960) and popularized by David Lewis's *Convention* (1969), emphasizing the importance of salience reasoning for solving coordination games.[1,2] The idea, roughly, is that, while these games

[1] E.g. Bicchieri (2005, p. 36ff), Cubitt & Sugden (2003), Gauthier (1975), Gilbert (1989), Postema (2008), Hédoin (2014), Lewis (1969) and Schelling (1960).

[2] The games that I will be talking about are two-player, conflict-free, pure coordination games, i.e. games with multiple strict Nash equilibria in which one player's gain does not require the other player's sacrifice. In this game, players have to solve the equilibrium selection problem. There are two relevant pure equilibria and and the players have to figure out a way to settle on one of them. Ultimately, each player is trying to match what she takes the other player to choose, which is why each player's choice depends only on estimates (beliefs, credences, or knowledge) about the other player's choice.

✉ Julius Schönherr
schoenherrjulius@gmail.com

1 Institute of Foreign Philosophy, Department of Philosophy and Religious Studies, Peking University, 5 Yiheyuan Rd., Office 223, Haidian Dist., Beijing 100871, China

Published online: 03 March 2021

have multiple Nash equilibria and can't be solved on apriori grounds, the fact that one equilibrium is *salient* can[3] explain why agents nevertheless manage to reliably solve recurring coordination problems.[4]

Among the various philosophical analyses of salience reasoning in coordination games, one has been especially pervasive.[5] According to this line of reasoning (e.g. Cubitt & Sugden, 2003; Hédoin, 2014; Sillari, 2005; Vanderschraaf & Sillari, 2014), Lewis's analysis as stated in *Convention* (1969) is basically right but can be helped by a more rigorous way of formalizing the relevant ideas.

As we shall see shortly, the Lewisian solution to coordination problems supplements salience-based inferences with higher-order beliefs, e.g. with beliefs about what the other player believes oneself will choose. It is this supplement that I shall attack in this paper: salience-based inferences are rationally impermissible when supplemented with such higher-order predictions. The arguments presented here are targeted specifically against the popular Lewisian model of salience-based coordination. This is done to anchor the discussion in the philosophical literature. The results, however, quite naturally generalize to any model that relies on such a supplement in an attempt to ground successful coordination in salience reasoning.

Without prematurely delving into any formalism, the relevant consensus about the Lewisian model can be captured in the four points stated just below. The first two of these points will be most essential for the arguments presented in the present paper. If salience[6] guides coordination, then the following must be true:

(1) *Public Fact Condition.* There is a public fact[7]—e.g. a visual cue, or a strong enough precedent—on the basis of which an agent may predict future behavior.
(2) *Symmetric Reasoning Condition.*[8] Both players are, and they know that they are, symmetric reasoners with regard to the relevant propositions, i.e. each player

[3] In some cases, salience might impede cooperation (see Gilbert, 1989, p. 66f). Suppose, for instance, you and I stand to win a prize if we both push the same color button at roughly the same time. Unfortunately, we cannot communicate to coordinate our actions. The buttons available are: red, green, blue, and yellow. Suppose a public announcement is made saying that we have a red phobia and would never press red buttons for any reason. Surely, this announcement would make the red button salient, but it wouldn't help us coordinate to push the red color button.

[4] The salience of a particular outcome is a correlation device, which is why the salient outcome has sometimes been called the "correlated equilibrium". This notion was first introduced by Aumann (1974) and then used by Vanderschraaf (1995) in a discussion of Lewis's *Convention*.

[5] Some alternative suggestions are the following. Gauthier argues that salience can change the structure of the game. Salience-reasoning, thus, is not reasoning about pure coordination (Gauthier, 1975). Sugden (2003) argues that players can coordinate their actions by conceiving of them as a "team". Postema (2008) thinks comparing coordination to jazz improvisation can help us understand coordination.

[6] I should add a note on the use of the term "salience". We shall say that public events such as a visual cue, or precedent *make* a coordination equilibrium salient. If a particular coordination equilibrium has precedent, then it makes sense to say that this equilibrium is salient *because* it has precedent. Of course, this does not amount to a definition of salience, which I do not intend to provide.

[7] In the present context, all we need is an informal concept of publicity, meaning the relevant fact is "out in the open" between both agents. I don't wish to commit to a formal account of publicity involving common knowledge (see Paternotte 2011 for such a definition).

[8] These labels—'Symmetric Reasoning Condition' and 'Public Fact Condition'—were suggested by an anonymous reviewer.

knows that any (relevant) inference she draws will likewise be drawn by the other player.

(3) Assumptions (1) and (2) imply common knowledge (belief, or reason to believe) that both players will choose their part of a particular coordination equilibrium; a *common expectation* as we might call it.

(4) This common expectation (from step 3) provides rational players with sufficient reason to play their part in a coordination equilibrium.

As indicated, my negative argument will be that this line of reasoning is self-undermining, at least if these conditions are applied simultaneously.[9] More particularly, the inference towards future behavior stated in (1) is false if (2) is true. This is so because *knowledge (or belief) that both players are symmetric reasoners defeats behavioral predictions grounded in public events such as precedent.* My positive argument (section three) will be that condition (1) can be applied only if a reasoner withholds judgment about the other player's reasoning process, that is, only if she doesn't simultaneously apply condition (2). She may, however, in a further step reason about the other player's reasoning process, thus, subsequently applying condition (2).

Now, statement (1) comes in two versions. A player may either use facts such as precedent to predict the other player's future behavior, or, alternatively, her own future behavior. Both versions should be addressed separately. Let me, for now, simply provide clear formulations of both ideas:

$(1)^O$ There is a public fact—e.g. a visual cue, or a strong enough precedent—on the basis of which an agent may predict *the other player's*[10] future behavior.

$(2)^S$ There is a public fact—e.g. a visual cue, or a strong enough precedent—that an agent may use as a guide[11] for *her own* behavior.

These senses are not always clearly distinguished in the literature. An exception is Gilbert (1989) who decidedly argues for the latter, contra the former, formulation. For the time being, we will stick to version $(1)^O$ and come back to Gilbert's proposal at the end of section two.

While I shall argue for the claim concerning the self-undermining nature of this line of reasoning in section two, I will here simply illustrate all this—steps (1) through (4) as well as my critical assessment—, first, with a vignette and thereafter with a brief formalization:

[9] I will address the codicil "if these conditions are applied simultaneously" at the very end of this paper.

[10] For simplicity, we'll confine ourselves to two player games throughout this paper.

[11] The use of the term "guide" instead of "predict" is preferable for the following reason. Many philosophers hold that an agent cannot predict her own behavior based on previous actions while deliberating what to do. For a nice summary of this debate see Hájek (2016). Furthermore, Gilbert's claims, which I will focus on below, are not couched in terms of prediction. On her view, rational agents may sometimes decide to give in to an "urge" to act in accordance with the salient option.

> **Fast food**. You and I want to meet for lunch. We have two options: McDonald's or Wendy's. We don't care where we'll have lunch as long as we'll have lunch together. In the past, we've always gone to McDonald's.

In this case, 'McDonald's' has a strong precedent. I shall assume that this precedent is common knowledge, i.e. we commonly know that that's where we've gone. According to the above line of reasoning, I may use this precedent and infer where you will go for lunch. We've reached the end of (1). Provided that I believe that we are symmetric reasoners—see step (2)—, I assume that you likewise predict my behavior using precedent as your standard of inference. Under the assumption that we're symmetric reasoners, we can suitably iterate the inference from step (1) and thereby reach a common expectation that we shall go to McDonald's. Finally, this common expectation rationalizes that each of us intend to go to McDonald's. If we act accordingly, and manage to meet at McDonald's, we've solved the coordination problem.

My critical claim, as applied to the above vignette, is that I may not predict your behavior using precedent given that I know that you predict my behavior using precedent. The reason is simple. If I know that you use precedent in predicting my behavior, it is this higher-order expectation alone that rationalizes my prediction of your action. Precedent is thereby defeated, or *excluded* as we will say, as a predictor of your behavior.

Another way to put the same point is as follows. Inferring future behavior from public facts such as precedent is valid only by default, e.g. the fact that we've always done something in the past (say) does not necessitate that this is what we shall do. Of course, the *mere* fact that these inferences are defeasible is unproblematic as long as possible interfering factors are incidental. The problem, as I see it, is that salience reasoning conducted along the above line systematically generates its own defeater and is thereby self-undermining.[12] I shall defend this point in the next section.

In the literature, we can find various ways to formalize steps (1) through (4). In what follows, I will avail myself of Hédoin's (2014) set-theoretic formalization. For our purposes, Hédoin's formalization is admirably concise, and nothing will depend on the formal minutiae. Cubitt and Sugden's (2003) alternative model formalizes Lewis's somewhat idiosyncratic notions of "reason to believe" and "indication". The result is an impressive, but intricate, formal model whose discussion would take us beyond what's necessary for this paper.

Now, let A be a proposition which is true at a world $\omega \in \Omega$ and let P designate a population with *i, j*, … members. Let X be a proposition whose content states a behavioral prediction of future behavior. Let K stand for the knowledge operator. To represent that a person *i* knows a proposition, we'll use subscript and write K_i. A might be the proposition "We've gone to McDonald's in the past" and X the proposition "The other player will go to McDonald's tomorrow". Proposition A refers to a *public event* (see statement (1) above) if, and only if,

[12] Weatherson (2016, section 1) presents a different argument for the seemingly self-undermining nature of higher-order reasoning in the context of coordination games.

$1a.\quad \omega \in K_i(A)$ each person i knows (has reason to believe) that A is true in the current state of the world.
$1b.\quad K_i(A) \subseteq K_i\left[K_j(A)\right]$ if i knows that A is true, then she knows that j knows that A is true.

Next, the idea that the agent infers future behavior based on A is captured by the following statement:

$1c.\quad K_i(A) \subseteq K_i(X)$ i infers X from her knowledge of A.

These statements formalize what's stated in (1) above. Lewis himself states all three conditions quite explicitly. "You and I have reason to believe that A holds. A indicates to both of us that you and I have reason to believe that A holds. A indicates to both of us that you will return" (Lewis, 1969, p. 52).

Lewis also recognizes the symmetric reasoner assumption: "You and I do have reason to believe we share the same inductive standards and background information, at least nearly enough so that *A* will indicate the same things to both of us." (Lewis, 1969, p. 53) This assumption, stated in (2) above, can be formally captured as follows:

$2a.\quad K_i(A) \wedge K_i\left[K_j(A)\right] \subseteq K_i\left[K_j(X)\right]$ given that i knows that j knows that A is the case, she knows that j knows that X is the case.[13]

These assumptions taken together imply the iterative chain of "*i* knows that *j* knows that… knows that X"; a common expectation, as we've called it earlier. A, in this case, is usually called a *reflexive common indicator* of X (e.g. Cubitt & Sugden, 2003; Hédoin, 2014, p. 370; Vanderschraaf & Sillari, 2014).

Critics of this (and related) model(s) have often questioned the plausibility of relying on *infinite* (or heavily iterated) chains of nested of higher-order expectations in the context of coordination games. Gilbert, for instance, famously argued that such infinite nesting merely yields conditional solutions of the following form: Player A chooses an outcome if she expects player B to choose it, and she expects B to choose it if she expects B to expect A to choose it, and so on *ad inf* (see Gilbert, 1989, p. 324).[14] Moreover, Lederman (2018a) argued that, given a *large* number of nested beliefs, the possibility of small mistakes and inaccuracies at each level in the cascade of reciprocal reasoning can undermine coordination. In contrast, the proposal presented in this paper targets *any kind* of higher-order reasoning. Salience-based solutions are principally incompatible with higher-order predictions of behavior. This concludes my rendition of salience-based reasoning as described

[13] Vanderschraaf and Sillari (2014) present the following alternative formulation: "Given a set of agents N and a proposition A′ ⊆ Ω, the agents of N are symmetric reasoners with respect to A′ (or A′-symmetric reasoners) iff, for each i, j ∈ N and for any proposition E ⊆ Ω, if Ki(A′) ⊆ Ki(E) and Ki(A′) ⊆ KjKj(A′), then Ki(A′) ⊆ KjKj(E).

[14] Consult Sillari (2005, p. 383) for a brief discussion and disagreement with Gilbert.

in the literature. Before I continue with arguments, let me provide a few notes of clarification.

1.1 Logic -v- probability-based approaches

There has been a parallel debate in economics—inaugurated by Rubinstein's (1989) remarkable paper "The electronic mail game: Strategic behavior under 'almost common knowledge'"—that likewise investigates the necessity of common knowledge in the context of coordination games.[15] Rubinstein-inspired discussions of the problem at hand are stated probabilistically involving credences. My presentation, on the other hand, follows a logic-based approach. While I do not doubt that the thoughts presented in this paper do generalize to probabilistic approaches, actually arguing that they do would be beyond the scope of my paper. Whether the notion of "defeat" that I rely on has an analogue in probabilistic approaches to belief is a matter of ongoing debate (e.g. Horty, 2012) which I can't hope to take on *en passant* as it were.

1.2 Idealization

The second note concerns various idealizing assumptions that I shall make. Throughout this paper, I shall assume common knowledge of rationality[16] and unbounded reasoning. These are strong assumptions to make and, therefore, in need of a bit of justification. First, these assumptions have often been relaxed to argue for a *particular* solution to philosophical problems about coordination. For instance, Harvey Lederman (2018c) has argued that coordination can be facilitated by dropping the assumption that the players' rationality is common knowledge. Although the players may be ideally rational, they might nevertheless not know that they are. Others (e.g. Kneeland, 2012; Schönherr, 2019) have argued that bounded reasoners can coordinate without common knowledge. Yet others (e.g. Skyrms, 2004) have explored successful coordination in entirely non-strategic contexts. In this paper, I won't be concerned with these solutions. Rather, my task is to explore a quite general defeat relation that traditional models of salience reasoning instantiate. For this reason, these idealizations should be taken as means to simplify the discussion, rather than substantive commitments. Furthermore, the claim presented in this paper is that salience-based predictions of behavior are incompatible with *any* (*i.e. even second-order*) higher-order behavioral predictions. Friends of the bounded rationality assumption, however, usually grant that players may reason through a couple of levels of higher-order reasoning before further nesting would overload their reasoning capacities (Clark, 1996, p. 95f).

[15] Others are Fagin et al. (1995, Ch. 6, Ch. 11), Fagin et al. (1999) and Halpern & Moses (1990).

[16] By "rational" I mean that (a.) the players seek to maximize individual expected utility, and (b) they know all propositions that can be derived in the context of the game.

1.3 Sources of salience

The example used throughout this paper is that of precedent as a source of salience. My arguments, however, apply to salience-based solutions more generally, insofar as they conform to the structure laid out above (see claims (1)–(4)). Suppose, for instance, that I promise to meet you at McDonald's for lunch tomorrow. We might think that my promise simply makes 'McDonald's' the salient solution, in which case we could simply interpret my promise to you as the value assumed by the variable A, indicating that my promise is a public event that can be used as a ground for salience reasoning. Alternatively, we might think that promises transform the nature of the game that we're involved in. For instance, my promising that I will show at McDonald's might create an additional normative obligation on my part that would persist even if I knew that you weren't going to show. If promises transform the game in this way, then my analysis does not apply, of course.

In the next section, I will argue that salience reasoning as presented in steps (1)–(4) is self-undermining. In section three, I will sketch a positive picture describing how salience reasoning may be permissibly conducted. The idea, roughly, is that we should think of salience reasoners as being *suspended* with regard to higher-order behavioral predictions.

2 Higher-order defeat

Let's remind ourselves of the first two conditions of salience reasoning:

> $(1)^{O}$ *Public Fact Condition.* There is a public fact—e.g. a visual cue, or a strong enough precedent—on the basis of which an agent may predict *the other player's* future behavior.
> (2) *Symmetric Reasoning Condition.* Both players are, and they know that they are, symmetric reasoners with regard to the relevant propositions, i.e. each player knows that any (relevant) inference she draws, will likewise be drawn by the other player.

Both claims taken together, as I shall argue in this section, are self-undermining if they are applied simultaneously. According to this model, when a player predicts how the other will act using, say, precedent as a standard of inference, she also assumes that the other player will use the same standard, i.e. that they reason symmetrically. The assumption, however, that the other player also uses precedent as her standard of inference makes one's own precedent-based inference impermissible in the first place. It "excludes" this inference as we shall say. In short, jointly applying condition (1) and (2) undermines the application of condition (1). Let's substantiate this idea.

The main intuition is that both assumptions taken together entail a higher-order behavioral expectation that defeats the initial inference that was stated in (1). The notion that statements (1) and (2) do in fact entail a higher-order expectation is a simple consequence of our definitions (1a–2a) and does not need further argument.

Thus, the philosophical work that will keep us occupied throughout this section consists in showing that these higher-order expectations do in fact act as a defeater.

Higher-order beliefs (or expectations) take precedent-based reasons out of consideration; they *exclude* them.[17] With regard to situations such as 'Fast Food' this means that once I have a higher-order expectation about where you think that I'll go, it is this higher-order expectation alone that justifies my first-order prediction about where I think you'll go.

The idea behind exclusionary defeat is sometimes illustrated using the following example: the fact that this chair looks red to me is a reason for thinking that it is red. If I learn, however, that the chair is illuminated by red light, then the fact that this chair looks red is no longer a reason for thinking that it is red. Thus, the fact that the chair is illuminated by red light takes the reason, constituted by the chair's seeming redness, out of consideration (e.g. Horty, 2012, p. 184; Pollock, 1970; Raz, 1975).[18],[19]

Exclusionary defeat should be distinguished from rebutting defeat. While an exclusionary defeater takes whatever is defeated out of consideration, rebutting defeat leaves the original reason intact; instead, a new reason of at least equal strength that favors an opposing conclusion is introduced. To see the contrast, consider also a case of rebutting defeat. The fact that Nixon is a Quaker is a reason for believing that he is a pacifist. And the fact that he is a republican is a reason for believing that he is not a pacifist. Even after you learn that he is a republican, the fact that he is a Quaker continues to be a reason for believing that he is a pacifist.

17 Defeaters of this kind have been called "exclusionary defeaters" (Horty 2012).

18 Horty's (2012) book *Reasons as Defaults* contains a user-friendly set-theoretic formal characterization of exclusionary defeat. The main idea is to supplement ordinary propositional logic with a special symbol '$\rightsquigarrow$' that represents defeasible generalizations. For instance, given two arbitrary propositions, *X* and Y, $X \rightsquigarrow Y$ stands for the defeasible generalization that lets a reasoner conclude Y from X, by default. Each such rule is denoted using a subscripted Greek letter δ_n (e.g. $\delta_1 = X \rightsquigarrow Y$). The conclusions of these default rules are picked out using the function $Con(\delta_n)$, e.g. $Con(\delta_1) = Y$. Exclusionary defeaters can be thought of as rules that take other rules out of consideration, they exclude them. Such rules are constructed using the special function $Out(\delta_n)$ which means that a rule is taken out of consideration and no conclusions may be derived from it. Reconsider the example of the illuminated chair. There are three relevant propositions: R = 'This chair looks red'; P = 'This chair is illuminated by red light'; F = 'This chair is red'. Furthermore, there are two relevant default rules: $\delta_1 = R \rightsquigarrow \mathrm{F}$ —this chair's seeming redness is a reason for concluding that it is red—, and $\delta_2 = P \rightsquigarrow \mathrm{Out}(\delta_1)$ —this chair's being illuminated by red light excludes the reason provided by this chair's seeming redness. For any given reasoning problem, we can collect these rules in a set *D*, e.g. for the chair illumination problem the set is this: $D = \{\delta_1 = R \rightsquigarrow \mathrm{F}, \delta_2 = P \rightsquigarrow \mathrm{Out}(\delta_1)\}$. To see which rules are excluded in our reasoning problem, we create a new set 'excluded(D)' that contains all rules that are taken out of consideration by some rule in *D*: excluded$(D) = \{\delta \in D : Con(D) \vdash Out(\delta)\}$. Now we have all the rules that we are *not* allowed to reason with. To obtain the rules we *are* allowed to reason with, we finally create a new set S that contains all and only those rules that are not excluded: S $= \{\delta \in D : \delta \notin excluded(D)\}$. This rendition is a vast simplification of Horty's model, but it should suffice to convey the basic idea.

19 The notion of defeat as employed in the present context is *normative*. If a reason is defeated, then it *ought not* to be used in drawing inferences. A reasoner may, of course, irrationally draw these inferences regardless. But if she does, her inference is normatively impermissible. Coordination among irrational players, as was indicated above, shall not be discussed in this paper.

Fig. 1 Rebutting defeat

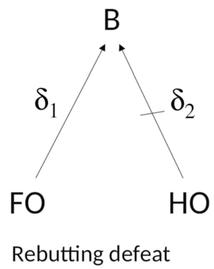
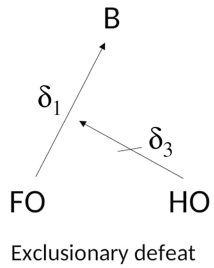


Fig. 2 Exclusionary defeat



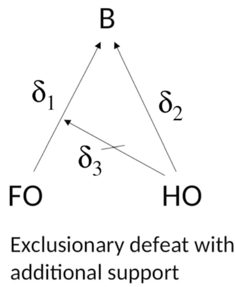
We can conveniently illustrate the contrast between both types of defeat using the inference graphs in Figs. 1 and 2 (see Horty, 2012, p. 24).

FO and HO represent reason-providing beliefs. In the context of the present discussion on coordination they represent the respective first-order and higher-order beliefs. Arrows represent the reasons that these beliefs provide. The arrows directed at B—in our context a prediction of behavior—represent the reason for B (regular arrow) and against B (strike-through arrow) respectively. δ_2 rebuts δ_1. δ_3 excludes δ_1 (Figs. 1 and 2).

I've pointed out that excluded considerations do not retain their supporting force when defeated. This is important. After all, the negative claim in this paper is that salience-based reasoning is self-undermining when it is based on both the 'Public Fact Condition' and the 'Symmetric Reasoning Condition'. The pivotal argument in favor of this idea is that higher-order expectations make precedent-based first-order predictions impermissible. If higher-order beliefs *exclude* precedent-based predictions, then this is straight-forward. After all, an excluded consideration is simply not available for reasoning. If, alternatively, this consideration was defeated by way of rebutting defeat, then the self-undermining nature of Lewis' model of coordination would not be immediately obvious.

We need to add one further element to complete our model of defeat. In coordination games, higher-order beliefs about what the other player thinks oneself will do not only exclude lower-order beliefs, they also support a behavioral prediction in their own right. In cases such as 'Fast Food', the higher-order prediction that you think I'll go to McDonald's gives me a reason to believe that that is where you'll go. This idea has yet to be incorporated into our model. Figure 3 provides and illustration of it.

Fig. 3 Exclusionary defeat with additional support



This illustration shows what we're after: higher-order expectations justify behavioral predictions while also excluding reasons that are potentially provided by lower-order behavioral predictions on the basis of, say, precedent.[20]

In the rest of this section, I will argue that this exclusionary defeat relation (indicated by δ_3) does in fact obtain. Consider an initial example:

> **Fast food 2.** You and I want to meet for lunch. We have two options: McDonald's or Wendy's. We don't care where we'll have lunch as long as we'll have lunch together. In the past, we've always gone to McDonald's.
> I learn that you believe that I will go to Wendy's this time around.

Given that I believe that you believe that I will go to Wendy's, I will expect you to try to match my choice. Hence, I expect you to go to Wendy's, which is why I myself ought to go to Wendy's. Given this set of beliefs, it shouldn't matter to me that we've always gone to McDonald's in the past. It seems that I may give precedent weight only if I thought that you weren't going to act on your beliefs; only if I thought you were blatantly irrational, that is.[21]

Let's examine a few more cases to expand on and strengthen my claim further. Consider a case in which I merely believe that you have *some*, albeit unspecified, belief about what I'm going to do. In this case, the defeat relation still obtains as the following vignette will make clear:

[20] There is a glaring similarity between the framework presented here and Muñoz's (2019) notion of "disqualification." One consideration, C1 disqualifies another consideration C2 if both considerations support the same conclusion, but the evidential support "comes from [the disqualifying consideration] alone" (Muñoz's, 2019, p. 888). Muñoz argues that disqualification is a *sui generis* relation that cannot be defined in terms of more mundane forms of defeat. Of course, here is not the place to decide the case. For our purposes, the more traditional notions of defeat (i.e. exclusionary, and rebutting defeat) have sufficient expressive power.

[21] Defeat relations among reasons can obtain for various reasons. In standard cases, more specific information defeats less specific information (e.g. Horty, 2012, p. 216). We wouldn't, for instance, want to conclude that Tweety can fly on the ground that he's a bird, knowing that he's penguin. One might, thus, wonder whether the defeat relation between precedent and higher-order beliefs can be explained in similar ways. I think this is not so. Rather, the defeat relation in our case is simply grounded in basic assumptions about the structure of the game. In pure coordination games, each rational player is trying to match the other's choice, i.e. each player will act on what she believes the other player is going to choose. If both the structure of the game as well as the players' rationality are common knowledge, then each player knows that the same holds true for the other player. Each player knows that the other will act on her belief about what she thinks the other is going to do, which is why precedent has, at this point, been defeated.

> **Fast food 3.** As before, you and I want to meet for lunch. We have two options: McDonald's or Wendy's. We don't care where we'll have lunch as long as we'll have lunch together. In the past, we've always gone to McDonald's.
> I learn that a source, who you (perhaps falsely) believe to be infallible, flipped a fair coin, and either told you that I would be at Wendy's this time (if it came up heads), or she told you that I would be at McDonald's (if it came up tails).

In this case, my rational response is debatable (we'll come back to this in the next section). One thing is clear, however: I should not rely on precedent in making my decision. You have a belief about where I'm going to be, and you will try to match my predicted choice accordingly. This way of reasoning about you is conclusive for me and precedent should therefore not be invoked. Thus, it is simply the fact that you have *some* higher-order belief that does the defeating.

Consider next a case in which I'm uncertain whether you have any higher-order expectation. In this case, precedent may intuitively be invoked in generating a prediction precisely to the extent to which I am convinced that you lack higher-order expectations about my behavior:

> **Fast food 4.** As before, you and I want to meet for lunch. We have two options: McDonald's or Wendy's. We don't care where we'll have lunch as long as we'll have lunch together. In the past, we've always gone to McDonald's.
> I learn that a source, who you (perhaps falsely) believe to be infallible, flipped a fair coin, and either told you that I would be at Wendy's this time (if it came up heads), or she didn't tell you anything at all (if it came up tails).

In 'Fast Food 4', the intuition that precedent may permissibly be invoked starts to make intuitive sense and is also supported by the thesis under discussion. In the context of the example, I assume that, in the event that your source didn't tell you anything at all, you have no belief about what I'm going to do, which is why precedent reemerges as a permissible source of inference.

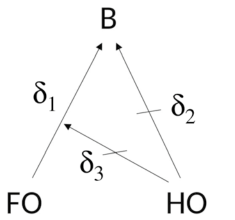
The cases discussed so far present the reader with situations in which two types of evidence *conflict*, i.e. cases in which precedent favors one response and higher-order beliefs favor a different response (see Fig. 4).

According to the Lewisian model of coordination, however, first-order and higher-order predictions support the same conclusion (see Fig. 3).

Concomitantly, what needs to be argued is that the higher-order beliefs continue to exclude lower-order beliefs even if both beliefs support the *same* behavioral prediction.

Now, δ_3 represents an exclusionary defeater. It attacks δ_1. On pain of restating the obvious, whether, beyond defeating δ_1, HO is also a reason in favor of B (see δ_2 in Fig. 3), in disfavor of B (see δ_2 in Fig. 4), or plainly neutral is simply a separate question. One thing that HO does is excluding δ_1; a *different* thing that HO might do is directly bear on B. Thus, based on these structural considerations, precedent can be excluded even if I learn that you've also used precedent to predict my behavior. If I learn that you predict my behavior using precedent, I'm decidedly *not* using precedent to predict your behavior even if both inferences support the same prediction.

Fig. 4 Exclusionary defeat and rebutting defeat



To see all this even more clearly, consider the following case:

> **Fast food 5.** As before, you and I want to meet for lunch. We have two options: McDonald's or Wendy's. We don't care where we'll have lunch as long as we'll have lunch together. In the past, we've always gone to McDonald's.
> First, a source, who I (perhaps falsely) believe to be infallible, tells me the following disjunction: either you predicted my behavior using precedent, in which case you'd be on your way to McDonald's; or you thought that I wanted my favorite burger, the Baconator as it turns out, in which case you predicted that I'd go Wendy's. Next, the source tells me which disjunct is true.

Precedent is defeated in the conflict (i.e. Wendy's) case, at least if what was argued so far is true. If, alternatively, my source tells me that you too went with precedent, I don't suddenly activate precedent as a standard of inference. The way I reason is the same either way: I learn how you reason, and that alone tells me where you will go, thus, defeating precedent as a source of inference. I won't, for example, start weighing both pieces of evidence—precedent-based and higher-order evidence—thinking that precedent is weaker in the situation at hand. Rather, precedent becomes a *non-issue*; it simple doesn't matter.

These vignettes (i.e. Fast Food 2 through 5) illustrate, each in their own way, the intended notion of higher-order defeat. In each of these examples, however, the higher-order belief is generated through an idiosyncratic process (e.g. a credible source flipping a coin). To be sure, these processes differ from the Lewisian framework as captured in the 'Symmetric Reasoning Condition':

(2) *Symmetric Reasoning Condition.* Both players are, and they know that they are, symmetric reasoners with regard to the relevant propositions, i.e. each player knows that any (relevant) inference she draws, will likewise be drawn by the other player.

Thus, whereas according to the Lewisian framework a reasoner generates a higher-order expectation using the 'Symmetric Reasoning Condition', the higher-order expectations in the above vignettes are generated through some alternative process. This disanalogy may look worrisome, and the suspicion may arise that we can't learn much from these vignettes. I think this suspicion is misplaced.

The extra information provided in the above vignettes is incidental. What matters in all of these examples is *the fact that* I expect you to have a belief about what I'm going to do. I don't care about how exactly you came to have this belief, e.g. whether a source told you where I'd go, whether I learn that you reason from precedent, whether you simply woke up with a higher-order belief on your mind, etc. The result is always the same: there is some process by which you acquired a belief. The process ultimately doesn't matter to me. What matters to me is that you have a belief about where I'll go. After I learn that you have such a belief, precedent becomes otiose as a reason for predicting your behavior. And in this crucial aspect, vignettes 2–5 cohere with the Lewisian framework of coordination.

Finally, according to the Lewisian model of reasoning in coordination games, higher-order and first-order predictions are not independent. On this model, I believe that you use precedent in trying to coordinate with me *precisely because I use it.* In this sense, some will think that, in the context of the Lewisian framework, higher-order beliefs cannot defeat precedent-based reasons, because, after all, such higher-order predictions themselves depend on precedent being used.[22] By way of reiterating what was just said, it is *the simple fact that* I have a higher-order expectation that does the defeating. How I came to have such an expectation is irrelevant; maybe someone informed me about the way you reason, maybe I know you and thus know how you tend to reason, or maybe I assume that you reason just the way I reason. It doesn't matter. In each of these cases, my higher-order expectation defeats my initial prediction that would be based on precedent.[23]

Now, I have attacked a pervasive line of reasoning according to which each player predicts *the other player's* behavior using precedent as a standard of inference. Alternatively, we may suppose that precedent somehow bears on the agent's *own* choices in a coordination game. This is Gilbert's (1989, p. 74) attempted solution:

> In the model I am now considering the main force operating in each case is not the agent's reasoning (about others or anything else) but his own unreasoned impulse. Reasonably ascribing a similar impulse to others (either through

[22] Precedent-based considerations may, as one helpful reviewer put it, seem to be "smuggled into" the higher-order predictions.

[23] Now, a critic may accept the idea that higher-order beliefs defeat precedent-based inferences, but she may, however, wish to add that these higher-order beliefs can themselves be defeated, in which case the original precedent-based inference would be reinstated. More particularly, the critic may wish to propose the following rule that governs cases of conflict between higher-order and lower order predictions: in such conflict cases, the lower-order reason or rule takes priority. This rule functions as an *arbiter*, as it were. *Formally*, this is a possibility, of course. Horty's (2012, 129) framework, which I have relied on throughout this paper, explicitly allows for the possibility that defeaters might themselves be defeated. In the case at hand, however, such an extra rule is not plausible. Reconsider a case akin to 'Fast Food 3'. Suppose I know two things. I know that we've always gone to McDonald's in the past. I also happen to know that you think that I will go to Wendy's this time around. If my arguments are on track, then my higher-order belief defeats my lower-order belief. According to the suggestion that there is a new 'arbiter' rule, my precedent-based prediction would win out. But this suggestion clearly delivers the wrong result. In this case, I ought to go to Wendy's. If I rely on precedent and, thus, go to McDonald's, I'm simply irrational.

observation of their actions or simply through knowledge that they are the same kind of creature) he sees no reason to struggle against this tendency.

The proposed solution appears towards the end of Gilbert (1989). In coordination games with a salient choice, Gilbert argues, agents might simply have an impulse to act in accordance with the salient option. Furthermore, a player is rationally permitted to give in to the impulse if she believes that the other player feels a similar impulse. Gilbert's idea maps onto our model stated in the introduction. She assumes a public event A—e.g. precedent or an announcement—which each player takes as a guide towards future behavior. The only difference is that facts such as precedent are not taken to bear on a prediction of the *other* player's future behavior but are rather taken to be a guide towards one's *own* future behavior. For this reason, I restated statement (1) in the introduction in order to accommodate Gilbert's thoughts:

(1)^{S} There is a public fact—e.g. a visual cue, or a strong enough precedent—that an agent may use as a guide for her own behavior.

Additionally, each player assumes that the other player is symmetrically impulsive (1989, 74).

Given everything I've said about higher-order defeat, I am in a position to refute Gilbert's proposal. To see how, consider again a lunch situation similar to the ones above. We've always gone to McDonald's. Suppose I have an impulse, grounded in precedent, to go to McDonald's. Suppose I learn that you will go to Wendy's this time around. Surely, I should not follow my impulse. I should go to Wendy's instead. My impulse as a guide for action has been defeated if I am at all rational. Consider next the reverse situation. I know that you've kept your impulse to go to McDonald's, but I've lost it. Still, if I'm rational, I will go to McDonald's. Thus, in both scenarios it is only my belief about your impulse that matters. At least in these conflict scenarios, my impulse as a guide for action is otiose. Above, however, I have argued that conflict is not necessary for defeat. Even if I learn that you go to McDonald's this time around, it is *not my impulse that I'm following*. Rather, it is my prediction about your behavior that does the work. This remains true, even if my impulse and my belief about your impulse line up, this latter belief defeats my impulse as a viable guide to action.

Now, I've argued that the presence of a higher-order belief defeats first-order salience-based inferences towards future behavior. Let me stress again why this makes salience reasoning, if construed along Lewisian lines, self-undermining. When trying to coordinate with you, I perform an inference whose conclusion is a piece of future behavior. If I furthermore assume that we're symmetric reasoners, I conclude that you perform the same inference. Unfortunately, believing that you perform this inference makes it inappropriate for me to perform this inference in the first place. However, if I may not perform it, then you, *qua* symmetric reasoner, will not perform it either.

The conclusion reached in this section might strike some readers as absurd. "Surely", they may think, "it is possible to successfully coordinate even if we know that the other player reasons as we do." Initially, one may even think that the symmetric reasoning condition serves as an enabling (not as a defeating) condition, as it were. The intuition is that one may rely on precedent in predicting the other's behavior only if it is presupposed that the other player does so as well. The current analysis contradicts this intuition.

While I shall use the next section to provide a solution, let me here add a general note on philosophical puzzles surrounding coordination. Over the years, philosophers have formulated a host of puzzles whose surprising conclusion is that, given only a few rather innocuous assumptions, coordination among rational agents seems to break down. And yet, *real* agents find it quite easy to coordinate their behavior in similar circumstances. To give just one example, according to the famous 'coordinated attack' problem, two rational generals, whose communication is confined to a less than perfectly reliable mailing system, fail to coordinate their attack if fewer than infinitely many messages are exchanged.[24]

Although in real life, people find it quite easy to coordinate their behavior in the relevant situations, it would surely be misguided to conclude that these theoretical puzzles and paradoxes must therefore be ill-described to begin with. Rather, we should see these paradoxes as an invitation to adjust, rethink, and fine-grain the assumptions that give rise to them. The problem I have described fits this tradition of paradoxes of coordination in that seemingly harmless assumptions lead to seemingly paradoxical conclusions when applied to a context of strategic coordination. And yet, just as in many other cases, the fact that real-life coordination seems unproblematic should not be seen as a reason to dismiss the proposal under discussion. Instead, the ease of real-life coordination is an invitation to refine our theory to account for it.

3 Salience reasoning and belief suspension

So far, this paper has been critical in nature. I've argued that an influential analysis underlying salience-based solutions to coordination games is self-undermining. In this section, I shall make a positive suggestion about how to fix it. I shall stick to precedent-based reasoning in stating and discussing my claims.

It is worth noting that a real fix is needed. Simply biting the bullet—i.e. denying that salience-based reasoning is ever appropriate—is a non-starter. Clearly, salience-reasoning often does help us coordinate, and it would be strange if successful, and non-accidental, coordination was to *require* some form of irrationality. I think we should aim to describe a situation in which ideal agents may permissibly invoke salience-based facts such as precedent in predicting the other player's behavior in the pursuit of solving a coordination problem.

To see how this would go, let's go back to 'Fast Food'. When thinking about which restaurant you will attend, I might simply not have formed any beliefs yet, i.e.

[24] Furthermore, Lederman (2018a) argues that, given standard assumptions, rational players fail to coordinate their efforts in the following situation: to win a large prize, two players, who cannot communicate, must each hit a buzzer if a mast they both clearly see right in front of them is larger than 100 cm. The mast is 300 cm high.

I might be *suspended* about the matter. Belief suspension about a proposition (say, P) entails, as almost everybody agrees, neither believing nor disbelieving that P is the case (e.g. Bergmann, 2005, p. 420; Wedgwood, 2002).[25] Of course, *simply* not believing and disbelieving is not sufficient for belief suspension. After all, a person who has never even considered a certain proposition is not suspended about it. She simply doesn't entertain this proposition. Belief suspension requires at the very least some form of cognitive contact with the pertinent proposition. It is, as Scott Sturgeon puts it, a state of "committed neutrality" (Sturgeon, 2010, p. 90). The exact form of cognitive contact is controversial but luckily unimportant for our purposes. It has been said that suspending requires "refraining'' (Moore, 1979), "withholding'', or "resisting'' (see Friedman, 2013 for a summary) believing. This is obviously not the place to adjudicate between these issues; however, I think it is important to keep in mind that an agent may consider a proposition and yet neither believe nor disbelieve it.

I will, however, assume that belief suspension does not *force* certain credences on a reasoner, e.g. credence 0.5, as the Principle of Indifference suggests. My discussion of 'Fast Food 3' shows why this is important. There, I argued that using precedent as a standard of inference is incompatible with such equally distributed credences. I can't, of course, hope to defend this understanding of belief suspension and will, therefore, simply have to assume that belief suspension does not force credences. Whether this is true is a matter of continuous debate.[26] In this sense, the idea presented in this section falls short of a full defense and should be regarded as an attractive (as I hope) how-possibly solution to the problem at hand.

Is there a candidate situation in which rational coordinating agents are suspended about the respective other's belief concerning her own choice; about the other player's belief about what oneself believes, the other player believes etc.? I think there is. This is the situation in which both agents *have just started deliberating about what to do*.

At the start of their deliberation, the agents haven't considered any evidence yet, which is why they should suspend belief about how the other player will act, what she believes about one's own actions, and so on *ad inf.* There are two reasons in support of this thought. First, it seems that if belief suspension is ever a rational response to one's epistemic situation, then this should be before any evidence has been considered. Jane Friedman emphatically states that "it is hard to think of evidential circumstances more appropriate for suspension […] The absence of evidence norm is among the most minimal ways for a subject to respect her evidence. It says just […] that when she has none, she may suspend." (Friedman, 2013, p. 61) Second, it is plausible to think that belief suspension is appropriate in deliberative contexts. This is the position recently defended by Friedman (2017) who states:

[25] Friedman has reservations (see Friedman, 2017). The connection between suspension and not believing, she contends, is *normative,* not descriptive. An agent who is suspended about P *ought not* believe nor disbelieve it; but since we're operating under the assumption of perfect rationality, we can sidestep these subtleties.

[26] E.g. Friedman (2013), van Fraassen (1998) and Hájek (1998).

> [W]e can say that there is nothing more to "opening a question in thought" than simply suspending judgment on that question. In suspending about *Q* we make *Q* an object of inquiry. From there we can wonder or be curious or deliberate (and so on) about *Q*. Suspending about a question puts that question on our research agenda. (Friedman, 2017, p. 26)

Deliberating, or "inquiring", about whether *Q* is true is most appropriate when we haven't settled on either believing or disbelieving it. In fact, according to Friedman, deliberation is the kind of activity that aims at resolving this neutral state of belief suspension.

Suppose that, in deliberating about how the respective other will act, we're initially suspended, because we haven't considered any evidence, and, as a corollary, have not formed any higher-order belief bearing on the other's choice. Suppose next that we (commonly) know that we've just started our deliberative process, and, thus, (commonly) know that we are so suspended. In this case, we haven't formed any higher-order beliefs about where the other thinks oneself will go. I don't have any belief about what you think I will do, and I also know that you don't have any such belief about what I think you will do. And because all potentially defeating higher-order beliefs are absent, we may, at least as far as the relevant defeaters go, permissibly predict the respective other's behavior using precedent as a standard of inference.

The following picture emerges: It is epistemically permissible for agents to predict each other's behavior based on precedent only in the absence of higher-order beliefs about their actions. The latter condition is satisfied (for instance) when the agents know that they've just started deliberating about how to act and are thus suspended about what the other thinks oneself will do, what she thinks oneself thinks the other will do etc.[27] Taking these considerations into account, we can replace the original 'Symmetric Reasoning Condition' with the following amendment:

> (2)* *Symmetric Suspension Condition.* Both players are, and they know that they are, suspended about whether they are symmetric reasoners with regard to the relevant propositions, i.e. each player is suspended about whether any (relevant) inference she draws will likewise be drawn by the other player.

Now, when stating the claim that Lewis-style reasoning in coordination games is self-undermining, I hedged and stated that such reasoning is self-undermining *if the pertinent conditions are applied simultaneously.* Thus, *in* predicting another player's behavior using precedent, a player needs to be appropriately suspended; but *after* having predicted another player's behavior using precedent, a player may reason

[27] Thus, salience-based coordination must, at its core, be vindicated without relying on higher-order reasoning requirements. This is not a bad thing, because higher-order belief requirements—paradigmatically, common knowledge requirements—are, as Lederman (2018c) points out, not usually meant to capture pre-theoretical desiderata. Rather, common knowledge requirements represent "simplifying technical assumptions" (Lederman 2018b). In fact, to many, common knowledge assumptions seem to be an implausible *departure* from common sense. In this sense, analyses that can vindicate rational cooperation without relying on higher-order reasoning models are, if anything, closer to common sense.

from higher-order beliefs. We can now see more clearly why this hedge was appropriate. After the coordinating parties have successfully employed precedent in predicting the other player's behavior,[28] they may, in a further step, assume that the other player has reasoned symmetrically. At this point, however, the precedent-based conclusion has already been drawn. Concomitantly, drawing this conclusion is not subject to exclusionary defeat. Although a reasoner may, thus, subsequently assume that they both reasoned symmetrically, the main upshot of the present paper remains true: *in* predicting a player's behavior using precedent as a standard of inference, a player may not assume that both reason symmetrically; rather, a player must be suspended about the other player's reasoning process. This idea was crystallized by replacing the initial 'Symmetric Reasoning Condition' with the 'Symmetric Suspension Condition.' Symmetric reasoning is, thus, not required to predict the other player's behavior using precedent, or other forms of salience, as a standard of inference.

Now, this paper has been about salience reasoning in *coordination games* quite generally. Let me, lastly, comment on how my arguments fit into the analysis of *conventions* quite specifically. After all, Lewis himself, whose model has been under discussion, focuses on conventions. Conventions (e.g. driving on the right side of the road) are highly complex phenomena, defined by a host of rather contentious features. Let me here confine myself to a single observation. Ultimately, Lewis simply builds higher-order expectations of behavior, implemented through a common knowledge requirement, right into the definition of a convention (Lewis, 1969, p. 76), e.g. a behavioral regularity is a convention only if all or most members expect each other to expect each other to follow it. If Lewis is right about this, and if what I've been arguing is correct, then the *maintenance* of an existing convention cannot involve precedent as a source of inference; after all, higher-order expectations exclude precedent as a source of inference. Precedent, in this sense, would play its part in *establishing* a convention. This is certainly in line with Lewis's observations and examples. Others (e.g. Cubitt & Sugden, 2003, p. 175; Sillari, 2008, p. 29) second the idea that different processes might be involved in establishing and maintaining conventions. It is worth mentioning, however, that the extent to which conventions require common knowledge is itself contentious (see Bicchieri, 2005, p. 38; Burge, 1975; Marmor, 2009 for some critical views). If it turns out that the problematic higher-order behavioral expectations are *not* constitutive of conventions, then this would seem to make room for the possibility that precedent-based reasoning may be important not just in establishing, but also in maintaining, conventions. Adjudicating this issue, however, will have to wait for some other time.

4 Conclusion

David Lewis wrote that salience (e.g. grounded in precedent) can support coordination by providing reasons for choosing a strategy when there is "no stronger ground for choice" (Lewis, 1969, p. 35). Higher-order predictions about what the

[28] Or their own behavior, according to Gilbert's (1989) line of reasoning that was addressed at the end of section 2.

other player thinks oneself will choose present, I have argued in this paper, such a "stronger ground for choice". For this reason, salience-based reasoning such as reasoning from precedent is legitimate only in the *absence of such higher-order behavioral predictions*. More concretely, I have pointed out that this absence requirement is satisfied when the agents commonly know that they both suspend belief about what the respective other is going to do and why she's going to do it. This claim is directed against a philosophical doctrine according to which salience reasoning must be supplemented with a symmetric reasoner assumption.

The idea that higher-order prediction requirements quite generally (and viz. common knowledge requirements quite specifically) should, in the context of coordination games, be couched in terms of *belief absences* or *belief suspension* has rarely been noticed. In fact, I only know of a few authors who have recognized such absences to be relevant in spelling out the conditions for coordination. Grice (1969, p. 159) suggests to regiment speaker meaning by *excluding* problematic higher-order intentions. In a discussion of "mutual" knowledge in communicative contexts, Martin Davies (1987, p. 717) suggests that "the philosophical work which was to be done by the notion of mutual knowledge should instead be assigned to a negatively characterized notion: the mutual absence of doubt." Third, Richard Moore (2013, p. 492), in his discussion of common knowledge in the context of conventional behavior, notes that "the extent to which common knowledge is necessary for conventional activity will be determined by its coordinative role. Such a role might consist in *protecting participants in a convention from higher-order doubts* about the conformity of others" (my italics). These somewhat cursory remarks merely hint at the structural importance that belief-absences have for solving coordination games. In this paper, I've elaborated on this idea. Importantly, the present analysis showed that higher-order beliefs—i.e. the constituents of common knowledge—are not simply an unnecessarily baroque theoretical element that clutters[29] our models of strategic reasoning in coordination games. Instead, their presence was shown to act as a defeater when trying to coordinate with others by relying on precedent or other sources of salience as a standard of inference.[30]

References

Aumann, R. (1974). Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics, 1,* 67–96.

Bergmann, M. (2005). Defeaters and higher-level requirements. *The Philosophical Quarterly, 55*(419–436), 8.

Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.

Burge, T. (1975). On knowledge and convention. *Philosophical Review, 84,* 249–255.

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Cubitt, R. P., & Sugden, R. (2003). Common knowledge, salience and convention: A reconstruction of David Lewis' game theory. *Economics and Philosophy, 19*(02), 175–210.

Davies, M. (1987). Relevance and mutual knowledge. *Behavioral and Brain Sciences, 10*(4), 716–717.

[29] This expression was suggested by a reviewer.

[30] I am immensely grateful to Peter Carruthers, Aleks Knoks, Harvey Lederman, Eric Pacuit, Javiera Perez Gomez, Arthur Schipper, and Aiden Woodcock for their comments on drafts of this paper. I am also grateful to Dominik Klein, and Olivier Roy for giving me the opportunity to present this work at the University of Bayreuth.

Fagin, R., Halpern, J. Y., Moses, Y., & Vardi, M. Y. (1995). *Reasoning about knowledge*. MIT Press.
Fagin, R., Halpern, J. Y., Moses, Y., & Vardi, M. Y. (1999). Common knowledge revisited. *Annals of Pure and Applied Logic, 96*(1–3), 89–105.
Friedman, J. (2013). Suspended judgment. *Philosophical Studies, 162*(2), 165–181.
Friedman, J. (2017). Why suspend judging? *Noûs, 51*(2), 302–326.
Gauthier, D. (1975). Coordination. *Dialogue, 14*(195–22), 1.
Gilbert, M. (1989). Rationality and salience. *Philosophical Studies, 57,* 61–77.
Grice, P. (1969). Utterer's meaning and intentions. *The Philosophical Review, 68,* 147–177.
Hájek, A. (2016). Deliberation welcomes prediction. *Episteme, 13*(4), 507–528.
Hájek, A. (1998). Agnosticism meets Bayesianism. *Analysis, 58*(3), 199–206.
Halpern, J. Y., & Moses, Y. (1990). Knowledge and common knowledge in a distributed environment. *Journal of the ACM (JACM), 37*(3), 549–587.
Hédoin, C. (2014). A framework for community-based salience: Common knowledge, common understanding and community membership. *Economics & Philosophy, 30*(3), 365–395.
Horty, J. F. (2012). *Reasons as defaults*. Oxford University Press.
Kneeland, T. (2012). *Coordination under limited depth of reasoning*. University of British Columbia Working Paper.
Lederman, H. (2018a). Uncommon knowledge. *Mind, 127*(508), 1069–1105.
Lederman, H. (2018b). Common knowledge. In M. Jankovic & K. Ludwig (Eds.), *Handbook of social intentionality*. London: Routledge.
Lederman, H. (2018c). Two paradoxes of common knowledge: Coordinated attack and electronic mail. *Noûs, 52,* 921–945.
Lewis, D. (1969). *Convention: A philosophical study*. Harvard University Press.
Marmor, A. (2009). *Social conventions*. Princeton University Press.
Moore, R. (2013). Imitation and conventional communication. *Biology and Philosophy, 28*(3), 481–500.
Moore, R. E. (1979). Refraining. *Philosophical Studies, 36*(407–24), 9.
Muñoz, D. (2019). Defeaters and disqualifiers. *Mind, 128*(511), 887–906.
Paternotte, C. (2011). Being realistic about common knowledge: A Lewisian approach. *Synthese, 183,* 249–276.
Pollock, J. (1970). The structure of epistemic justification. In *Studies in the theory of knowledge, American philosophical quarterly monograph series* (Vol. 4, pp. 62–78). Basil Blackwell Publisher, Inc..
Postema, G. J. (2008). Salience reasoning. *Topoi, 27*(1–2), 41–55.
Raz, J. (1975). Practical reasoning and norms. Hutchinson and Company, 1975. (Second edition) with new Postscript printed in 1990 by Princeton University Press, and reprinted by Oxford University Press in 2002; pagination refers to the Oxford edition.
Rubinstein, A. (1989). The electronic mail game: Strategic behavior under "almost common knowledge." *The American Economic Review, 79,* 385–391.
Schelling, T. (1960). *The strategy of conflict*. Harvard Business School Press.
Schönherr, J. (2019). Lucky joint action. *Philosophical Psychology, 32*(1), 123–142.
Sillari, G. (2005). A logical framework for convention. *Synthese, 147*(2), 379–400.
Sillari, G. (2008). Common knowledge and convention. *Topoi, 27*(1–2), 29–39.
Skyrms, B. (2004). *The stag hunt and the evolution of social structure*. Cambridge University Press.
Sturgeon, S. (2010). Confidence and coarse-grained attitudes. In *Oxford studies in epistemology* (Vol. 3, p. 21). Oxford: Oxford University Press.
Sugden, R. (2003). The logic of team reasoning. *Philosophical Explorations, 6*(3), 165–181.
Vanderschraaf, P. (1995). Convention as correlated equilibrium. *Erkenntnis, 42*(1), 65–87.
Vanderschraaf, P., & Sillari, G. (2014). Common knowledge. In E. N. Zalta (Ed.), *The Stanford encyclopaedia of philosophy*, Spring 2014 edn. Metaphysics Research Lab, Stanford University.
van Fraassen, B. C. (1998). The agnostic subtly probabilified. *Analysis, 58*(212–220), 3.
Weatherson, B. (2016). Games, beliefs and credences. *Philosophy and Phenomenological Research, 92,* 209–236.
Wedgwood, R. (2002). The aim of belief. *Philosophical Perspectives, 36*(s16), 267–297.