Rationality as conformity

Hosni, Hykel and Paris, Jeff

2005

MIMS EPrint: **2005.31**

Manchester Institute for Mathematical Sciences
School of Mathematics

The University of Manchester

Reports available from: http://eprints.maths.manchester.ac.uk/
And by contacting: The MIMS Secretary
School of Mathematics
The University of Manchester
Manchester, M13 9PL, UK

ISSN 1749-9097

Rationality as conformity*

Hykel Hosni† and Jeff Paris
Department of Mathematics, Manchester University, Manchester, UK.
{hykel,jeff}@maths.man.ac.uk

May 19, 2005

Abstract

We argue in favour of identifying one aspect of rational choice with the tendency to conform to the choice you expect another like-minded, but non-communicating, agent to make and study this idea in the very basic case where the choice is from a non-empty subset K of 2^A and no further structure or knowledge of A is assumed.

KEYWORDS: Rationality; Common Sense; Coordination Games; Uncertainty; Principle of Charity; Social Choice Theory; Reasons.

1 Introduction

The investigation described in this paper has its origins in [20, 21, 22] (see also [19] for a general overview). In those papers it was shown that as far as probabilistic uncertain reasoning is concerned there are a small set of so called 'common sense' principles which, if adhered to, completely determine any further assignment of beliefs, i.e. probabilities. Interesting as these results may be this raises the question *why* we consider these principles to be 'common sense' (or, more exactly, why we consider transgressing them to be *contra* common sense).

It is a question we have spent some effort trying to resolve. The principles looked to us like common sense, and indeed the general consensus of

*Preliminary draft of the paper published in *Knowledge, Rationality and Action.* Volume 144, Number 2, pp. 249-285, 2005.

†Partially supported by EPSRC research studentship and a grant from the Alexander von Humboldt Foundation, the Federal Ministry of Education and Research and the Program for the Investment in the Future (ZIP) of the German Government.

colleagues was that, certainly, to flout them was to display a lack of common sense. Nevertheless we could find no more basic element to which they could be reduced (for example showing as in the Dutch Book argument that if you fail to obey them then you are certain to lose that most basic of all substances, money). From this apparent impasse one explanation did however suggest itself. Namely, that these principles appeared common sensical to us all *exactly because their observance forced us to assign similar probabilities*. It is this idea, of common sense, or rationality, as conformity, that we shall investigate in this paper.

Certainly in the real world some one not acting in the way that people expect would be described as having *no common sense*, for example filling-up the home fridge with fresh food the day before leaving for a long holiday, or, in more serious situations, such as declaring war on your ally when already fully stretched, of acting *illogically* or *irrationally*. Despite the numerous meanings or even intuitions that have been attached to these terms, see for example the volume [6], for the limited purposes of this work we shall use them synonymously.

To motivate the sort of problem we are interested in suppose that your wife is coming to your office to collect the car keys but unexpectedly you have to go out before she arrives. Your problem is where to leave the keys so that she can find them. In other words your problem is *choosing* a point in the room where you think your wife will also choose to look. Being a logical sort of person you ask yourself "where would I expect someone to leave the keys?". If there was a vanity table by the door that might seem an obvious choice, because people tend to leave 'outdoor things' at this point. On the other hand if you had only just moved into the office and it only contained packing cases scattered around the walls then you might feel the centre of the carpet was the best option available to you, it being the only place, as far as you could see, that *stood out*.

It would seem in this situation that there are two considerations you could be drawing on. One is *common knowledge*, you assume that your wife is also aware of the typical use that vanity tables by entrances are put to. The other is what one might call *common reasoning*, you assume that your wife will also reason that the centre of the room 'stands out', so given the common intent to locate the same spot in the room, you place the keys right there. In the first case, conformity would be characterized as a consequence of learned and possibly arbitrary conventions. A formalization of this is not, however, what we are pursuing here. Indeed part of what we aim at understanding is how certain conventions might arise in the first place: why certain choices look more rational than others given that both agents intend to conform. So it is the second aspect of common sense –common reasoning– that we wish to

investigate in this paper.

To do this we shall take what might be described as a mathematician's approach to this problem. We shall strip away all the inessentials, all the additional considerations which one normally carries with one in problems such as the one described above[1], and consider a highly idealized and abstract simplification of the problem. Our justification for this is that if one cannot resolve this problem satisfactorily how could one expect to be successful on the infinitely more complicated real world examples?

2 The Problem

The problem we wish to consider is that of trying to choose one from a number of options so that your choice *conforms* with that of another like-minded, but otherwise inaccessible, agent (the payoff for success, ditto failure, being the same in all cases).

What is arguably the simplest possible choice situation of this sort is the one in which we have some finite non-empty set K of otherwise entirely structureless options f. In other words options that whilst different are otherwise entirely indistinguishable. Then the very definition of 'indistinguishable' seems to suggest that in this case there is no better strategy available to us than to make a choice from K entirely at random (i.e. according to the uniform distribution).

The inevitable next step then is to consider the case when we *do* have some structure on the options, or as we may henceforth call them, *worlds*, $f \in K$. In this case, as logicians, the most obvious minimal structure on these worlds is that there are some finite number of unary predicates which each of them may or may not satisfy. To simplify matters for the present we shall further assume that each world is uniquely determined by the predicates it does or does not satisfy. In other words we are moving up from the language of equality to a finite unary language. What this amounts to then is that K is a non-empty subset of 2^A, the set of maps f from the finite non-empty set A into $2 = \{0, 1\}$.

To give a concrete example of what is involved here we might have $A = 4$ and K the set of functions (worlds) $\{f_1, f_2, f_3, f_4, f_5\}$ where

[1]You have doubtless already though whyever doesn't he leave a message stuck to the door, or call her mobile, or leave the keys with his secretary, ... !

	0	1	2	3
f_1	0	0	0	1
f_2	0	1	0	0
f_3	0	1	1	0
f_4	1	1	1	1
f_5	0	0	1	0

and the problem, for an agent, is to pick one of these so as to agree with the choice made by another like-minded, but otherwise non-communicating and indeed, inaccessible, agent. However in presenting the problem like this we should be aware that as far as the agents are concerned there is not supposed to be any structure on A or $\{0,1\}$, nor even on K beyond the fact that it is the (unordered) set $\{f_1, f_2, f_3, f_4, f_5\}$. For practical examples this can be accomplished by informing the first agent that his or her counterpart may receive the matrix

$$\begin{array}{cccc} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{array}$$

with the columns permuted and the rows permuted.

We understand a non-empty subset K of 2^A as *knowledge*, indeed knowledge that among the elements of K only one of them corresponds to the world chosen by another like-minded agent facing the same choice. In this way we implicitly introduce a qualitative measure of uncertainty: the bigger the size of K, the greater the agent's uncertainty about which choice of worlds qualifies as rational. This corresponds to a very general and fundamental idea in the formalization of reasoning under uncertainty (see e.g. [7]) and plays a likewise important role here.

It is clear that in general there will be situations, as in the case where we assumed no structure at all, when the agent is reduced to making some purely random choices. We shall therefore assume that the agent acts by first applying some considerations to reduce the set of possible choices $K\,(\neq \emptyset)$ to a non-empty subset $R(K)$ of K and then picks at random from $R(K)$. A function

$$R : \wp^+(2^A) \longmapsto \wp^+(2^A),$$

where $\wp^+(2^A)$ is the set of non-empty subsets of 2^A (which for brevity we sometimes denote as $\mathbb{K}$), will be called a *Reason* if $R(K) \subseteq K$ for all $K \in \wp^+(2^A)$.

Clearly, then, an optimal reason R, is one that always returns a singleton $R(K)$ for all $K \in \wp^+(2^A)$, as this would amount to entail conformity with probability 1. We shall see, however, that this situation represents the exception rather than the rule in the formalization to follow.

One might question at this point whether a better model for the agent's actions might be to have him or her put a probability distribution over K and then pick according to that distribution. In fact in such a case the agent would do at least as well by instead selecting the most probable elements of K according to this distribution and then randomly (i.e. according to the uniform distribution) selecting from them – which puts us back into the original situation.

In the next three sections we consider three different Reasons which are suggested by the context of this, and related, problems.

3 The Regulative Reason

As mentioned already the work in this paper was in part motivated by considering why the principles of probabilistic uncertain reasoning introduced in [20, 21, 22] warranted the description 'common sense'. The underlying problem in those papers was analogous to the one we are considering here, how to sensibly choose one probability function out of a set of probability functions. The solution we developed there was not to directly specify a choice but instead to require that the choice process should satisfy these principles and see where that landed us. In fact it turned out well in the linear cases considered in [20, 21] since the imposed principles happily permitted only one possible choice.

Given that fortunate outcome there it would seem natural to attempt a similar procedure here, namely to specify certain 'common sense' principles we would wish the agent's Reasons to satisfy and see what comes out. Clearly, the present problem is much less structured then the one in which knowledge and belief are represented via subjective probability functions. Indeed the current setting is arguably one of the simplest ones in which we can make sense of rational choice concerning "knowledge" and "possibilities". It therefore follows that if choice processes analogous to the ones that characterize probabilistic common sense could be specified, those would have an undoubtedly high level of generality.

Our next step then is to introduce 'common sense principles' or rules that, arguably, Reasons *should* satisfy if they are to prevent agents from undertaking "unreasonable steps"[2]. Hence, we call the resulting Reason,

[2]See section 7 for more on this.

Regulative. The key result of this section is that their observance leads to a characterization of a set $R(K)$ of "naturally outstanding elements" of K, formulated in Theorem 1.

Renaming Let $K \in \mathbb{K}$ and let σ be a permutation of A. R satisfies *Renaming* if whenever

$$K\sigma = \{ f\sigma \mid f \in K \}$$

then $R(K\sigma) = R(K)\sigma$.

In this definition $K\sigma$ is, as usual, the set $\{ f\sigma \mid f \in K \}$, and similarly for $R(K)\sigma$ etc.. The justification for this seems evident given the discussion in the previous section. Since the elements of A have no further structure any permutation of these elements simply produces an exact replica of what we started with. More precisely if you feel that the most popular choices of worlds from K are the set of worlds $R(K)$ then you should feel the same for these replicas, i.e. that the most popular choices of worlds from $K\sigma$ should be $R(K)\sigma$.

Obstinacy

R satisfies *Obstinacy* if whenever $K_1, K_2 \in \mathbb{K}$ and $R(K_1) \cap K_2 \neq \emptyset$ then $R(K_1 \cap K_2) = R(K_1) \cap K_2$.

The justification for this principle is that if you feel the most popular choices in K_1 are $R(K_1)$ and some of these choices are in K_2 then such worlds will remain the most popular even when the choice is restricted to $K_1 \cap K_2$.

This 'justification' *in general* is more than a little suspect. For consider $f \in R(K_1) - K_2$. In that case one might imagine those agents who chose f from K_1 having to re-choose when K_1 was refined to $K_1 \cap K_2$. The assumption is that they went back to $R(K_1)$ and randomly chose from there an element which *was* in K_2. An argument against this is that by intersecting K_1 with K_2 some otherwise rather nondescript world from K_1 becomes, within $K_1 \cap K_2$, sufficiently distinguished to be a natural choice. Whilst this will become clearer later when we have other Reasons to hand it can nevertheless still be illustrated informally at this point.

Suppose that K is

$$\begin{array}{cccc} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{array}$$

In this case the two most obvious choices would appear (to most people at least) to be 0 0 0 0 and 1 1 1 1. However if we take instead the subset

$$\begin{array}{cccc} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{array}$$

of K then it would seem that now 1 1 0 0 has become the obvious choice, not 0 0 0 0 or 1 1 1 1.

Despite this shortcoming in certain cases we still feel is of some theoretical interest at least to persevere with this principle, and also because of the conclusions it leads to. We note that in so as far as nothing is known about the nature of the options, the property captured by Obstinacy is widely endorsed by the social choice community (see, e.g. [12]). Indeed there are also a number of related principles in that discipline which may warrant consideration vis-a-vis our present intention, though our initial investigations to date along these lines have not yielded any worthwhile new insights.

In order to introduce our final principle we need a little notation. For $K \in \mathbb{K}$ we say that $X \subseteq A$ is a *support* of K if whenever $f, g \in 2^A$ and f restricted to X (i.e. $f \restriction X$) agrees with g restricted to X then $f \in K$ if and only if $g \in K$.

The set A itself is trivially a support for every $K \in \mathbb{K}$. More significantly it is straightforward to show that the intersection of two supports of K is also a support, and hence that every $K \in \mathbb{K}$ has a unique smallest support. Notice that if K has support X then $K\sigma$ has support $\sigma^{-1}X$.

If K has support X then it is useful to think of this knowledge as telling the agent (just) how elements of K act on X. Namely, for f to be in K it is necessary and sufficient that $f \restriction X = g$ for some

$$g \in \{\, h \restriction X \mid h \in K \,\}.$$

Irrelevance

Suppose $K_1, K_2 \in \mathbb{K}$ with supports X_1, X_2 respectively and for any $f_1 \in K_1$ and $f_2 \in K_2$ there exists $f_3 \in \mathbb{W}$ such that $f_3 \restriction X_1 = f_1 \restriction X_1$ and $f_3 \restriction X_2 = f_2 \restriction X_2$. Then

$$R(K_1) \restriction X_1 = R(K_1 \cap K_2) \restriction X_1 \tag{Irr}$$

where

$$R(K) \restriction X = \{\, f \restriction X \mid f \in R(K) \}.$$

The condition on K_1, K_2 amounts to saying that *as far as K_1 is concerned K_2 is irrelevant (and conversely)* because given that we know (only) that f satisfies the requirement for membership of K_1 (i.e. that $f \restriction X_1$ is amongst some particular set of functions on X_1) the additional information that $f \in K_2$ tells us nothing we didn't already know about $f \restriction X_1$.

The principle then amounts to saying that in these circumstances the choices from K_1 and K_2 should also reflect that irrelevance. That is, if $f_1 \in R(K_1)$, then there is an $f_3 \in R(K_1 \cap K_2)$ such that $f_3 \restriction X_1 = f_1 \restriction X_1$ and conversely given $f_3 \in R(K_1 \cap K_2)$ there exists such a f_1 (and similarly for K_2).

The justification for this is along the following lines. In choosing a most popular point from K_1 we are effectively choosing from $K_1 \restriction X_1$ and then choosing from all possible extensions (in $\mathbb{W}$) of these maps to domain A, and similarly for K_2. The given conditions allow that in choosing from $K_1 \cap K_2$ we can first freely choose from $K_1 \restriction X_1$ then from $K_2 \restriction X_2$ and finally freely choose from all possible extensions to domain A. Viewed in this way it seems then that any function in $R(K_1) \restriction X_1$ should also be represented in $R(K_1 \cap K_2) \restriction X_1$[3].

Definition. *We shall say that a reason R is a* Regulative Reason *if is satisfies Renaming, Obstinacy and Irrelevance.*

3.1 The Regulative Reason characterized

We start by noticing that there certainly is one Reason satisfying the common sense properties defined above, namely the *trivial Reason R* such that $R(K) = K$ for all $K \in \mathbb{K}$, though of course in practice this 'reason' amounts to nothing at all[4].

Theorem 1. *Let R be a Regulative Reason. Then either R is trivial or $R = R_0$ or $R = R_1$ where for $i = 0, 1$ R_i is defined by*

$$R_i(K) = \{\, f \in K \mid \forall\, g \in K, |f^{-1}(i)| \geq |g^{-1}(i)| \,\}.$$

Conversely each of these three Reasons are Regulative, i.e. satisfy Renaming, Obstinacy and Irrelevance.

[3]There seems to be an implicit assumption in this argument that for $f \in K_1$, $f \restriction X_1$ and $f \restriction A - X_1$ are somehow independent of each other. In the current simple case of $\mathbb{W} = 2^A$ this is true but it fails in the case, not considered here, in which the worlds are probability functions.

[4]It can be shown that if we had taken A to be infinite then this would have been the only Regulative Reason.

We begin with the proof of the "if" part. As usual, $\vec{0} : A \to 2$ is defined by $\vec{0}(x) = 0$ for all $x \in A$ and similarly, $\vec{1} : A \to 2$ is defined by $\vec{1}(x) = 1$ for all $x \in A$.

The first step consists in showing that Regulative Reasons are indeed three-fold.

Lemma 2. *Let R be Regulative. Then either $R(2^A) = 2^A$ or $R(2^A) = \left\{\vec{0}\right\}$ or $R(2^A) = \left\{\vec{1}\right\}$.*

Proof.

We first show the following claim:

If $f, g \in R(2^A)$ (possibly $f = g$) are such that $0, 1$ are in the ranges of f, g respectively, then $R(2^A) = 2^A$.

To this end let $f, g \in R(2^A)$ and $f(x) = 0$ and $g(y) = 1$ for some $x, y \in A$. For σ a permutation on A transposing only x and y we have that $2^A\sigma = 2^A$. Hence, by Renaming, $R(2^A)\sigma = R(2^A\sigma)$. In particular:

$$f \in R(2^A) \Rightarrow f\sigma \in R(2^A). \tag{1}$$

Now let $K = \left\{h \in 2^A \mid h(y) = 0\right\}$. Since $f\sigma \in R(2^A) \cap K \neq \emptyset$ then:

$$\begin{aligned} R(2^A) \cap K &= R(2^A \cap K) \quad \text{(by Obstinacy)} \\ &= R(K). \\ \therefore f\sigma &\in R(K). \end{aligned} \tag{2}$$

Put $K_1 = 2^A$ and $K_2 = K$ with support $X_1 = A - \{y\}$ and $X_2 = \{y\}$, respectively. As $\emptyset = \{y\} \cap X_1$, we can, for any $f_1 \in 2^A$ and $f_2 \in K$, construct a function $f_3 \in 2^A$ such that $f_3 \restriction X_1 = f_1 \restriction X_1$ and $f_3 \restriction X_2 = f_2 \restriction X_2$. Thus

$$\begin{aligned} R(2^A) \restriction X_1 &= R(2^A \cap K) \restriction X_1 \quad \text{(by Irrelevance)} \\ &= R(K) \restriction X_1. \end{aligned} \tag{3}$$

Therefore, $g \restriction X_1 \in R(K) \restriction X_1$. Furthermore for

$$g'(z) = \begin{cases} g(z) & \text{if} \quad z \neq y \\ \\ 0 & \text{if} \quad z = y. \end{cases} \tag{4}$$

we have that $g' \in R(K)$. Hence $g' \in R(2^A)$, by (2) above.

The claim now follows since we have shown that if we take any function $h \in R(2^A)$ and change its value on one argument the resulting function is also in $R(2^A)$.

The proof of Lemma 2 now follows by noticing that if $R(2^A) \neq 2^A$ then by the claim either 0 or 1 is not in the range of any $f \in R(2^A)$. Therefore, since $R(2^A) \neq \emptyset$ it must either be that $R(2^A) = \{\vec{0}\}$ or $R(2^A) = \{\vec{1}\}$. ■

Our next step is to prove the required result for trivial Reasons.

Lemma 3. *If $R(2^A) = 2^A$, then $R(K) = K$ for any $K \in \mathbb{K}$.*

Proof. Notice that if $R(2^A) = 2^A$ then for $K \in \mathbb{K}$,

$$K \cap R(2^A) = K \neq \emptyset$$

so by Obstinacy,

$$R(K) = K \cap R(2^A) = K.$$

■

Hence, the final step in the proof of the "if" direction of Theorem 1 deals with the more interesting case of non-trivial Reasons.

It will be useful here to introduce a little notation. For the remainder of this section, let $\pi : dom(\pi) \to \{0, 1\}$, where the domain of π, $dom(\pi)$, is a subset of A. Similarly for $\pi_1, \ldots, \pi_k$. For such a π let

$$X_\pi = \{\, f \in 2^A \mid f \restriction dom(\pi) = \pi \,\}.$$

Lemma 4. *If $R(2^A) = \left\{\vec{1}\right\}$, then*

$$R(X_\pi) = \left\{\pi \vee \vec{1}\right\},$$

where

$$\pi \vee \vec{1}(x) = \begin{cases} \pi(x) & \text{if } x \in dom(\pi) \\ \\ \vec{1}(x) & \text{otherwise.} \end{cases} \tag{5}$$

Proof. Suppose that $z \in A - dom(\pi)$. To prove the result it is enough to show that $f(z) = 1$ for $f \in R(X_\pi)$. Let $K_1 = 2^A$ with support $\{z\}$ and $K_2 = X_\pi$ with support $dom(\pi)$. Notice that the conditions for the applications of Irrelevance are met since $\emptyset = \{z\} \cap dom(\pi)$. Hence

$$R(2^A) \restriction \{z\} = R(X_\pi) \restriction \{z\}.$$

Therefore, for $f \in R(X_\pi)$

$$f(z) = \begin{cases} 1 & \text{if } z \in A - dom(\pi), \\ \\ \pi(x) & \text{if } z \in dom(\pi), \end{cases} \tag{6}$$

making $f = \pi \vee \vec{1}$.

■

This can be immediately generalized as follows.

Lemma 5. *Suppose* $R(2^A) = \left\{\vec{1}\right\}$ *and let* $Z = \{z_1, z_2, \ldots, z_n\} \subseteq A$ *with* $0 \leq r \leq n$. *Let* $\tau_1^r, \tau_2^r, \ldots, \tau_q^r$ *be all the maps from a subset of size* r *of* Z *to* $\{0\}$. *Then*

$$R(X_{\tau_1^r} \cup X_{\tau_2^r} \cup \ldots \cup X_{\tau_q^r}) = \left\{\tau_1^r \vee \vec{1}, \tau_2^r \vee \vec{1}, \ldots, \tau_q^r \vee \vec{1}\right\}.$$

Proof.

We first recall that, by the definition of R,

$$R\left(X_{\tau_1^r} \cup X_{\tau_2^r} \cup \ldots \cup X_{\tau_q^r}\right) \subseteq (X_{\tau_1^r} \cup X_{\tau_2^r} \cup \ldots \cup X_{\tau_q^r}) \tag{7}$$

Now let π be a permutation of A such that $Z\pi = Z$. Then

$$\left(X_{\tau_1^r} \cup X_{\tau_2^r} \cup \ldots \cup X_{\tau_q^r}\right)\pi = \left(X_{\tau_1^r} \cup X_{\tau_2^r} \cup \ldots \cup X_{\tau_q^r}\right).$$

Hence, by Renaming:

$$f \in R\left(X_{\tau_1^r} \cup X_{\tau_2^r} \cup \ldots \cup X_{\tau_q^r}\right) \iff f\pi \in R\left(X_{\tau_1^r} \cup X_{\tau_2^r} \cup \ldots \cup X_{\tau_q^r}\right) \tag{8}$$

By equation (7), $R\left(X_{\tau_1^r} \cup X_{\tau_2^r} \cup \ldots \cup X_{\tau_q^r}\right) \cap X_{\tau_j^r} \neq 0$, for some $0 \leq j \leq q$. Thus, by Obstinacy,

$$\begin{aligned} R\left(X_{\tau_1^r} \cup X_{\tau_2^r} \cup \ldots \cup X_{\tau_q^r}\right) \cap X_{\tau_j^r} &= R\left(\left(X_{\tau_1^r} \cup X_{\tau_2^r} \cup \ldots \cup X_{\tau_q^r}\right) \cap X_{\tau_j^r}\right) \\ &= R\left(X_{\tau_j^r}\right) \quad \text{(for some } 0 \leq j \leq q\text{)}. \end{aligned} \tag{9}$$

Recalling, from Lemma 4, that $R\left(X_{\tau_j^r}\right) = \left\{\tau_j^r \vee \vec{1}\right\}$ we have that $\tau_j^r \vee \vec{1} \in R\left(X_{\tau_1^r} \cup X_{\tau_2^r} \cup \ldots \cup X_{\tau_q^r}\right)$ for some $0 \leq j \leq q$. By equation (8), however, this can be generalized to any $0 \leq j \leq q$. Hence

$$R\left(X_{\tau_1^r} \cup X_{\tau_2^r} \cup \ldots \cup X_{\tau_q^r}\right) \supseteq \left\{\tau_1^r \vee \vec{1}, \tau_2^r \vee \vec{1}, \ldots, \tau_q^r \vee \vec{1}\right\}. \tag{10}$$

To see that the converse is also true, suppose $h \in R\left(X_{\tau_1^r} \cup X_{\tau_2^r} \cup \ldots \cup X_{\tau_q^r}\right)$. Then since

$$R\left(X_{\tau_1^r} \cup X_{\tau_2^r} \cup \ldots \cup X_{\tau_q^r}\right) \subseteq X_{\tau_1^r} \cup X_{\tau_2^r} \cup \ldots \cup X_{\tau_q^r},$$

$h \in X_{\tau_j^r}$, for some j. But as we have just observed,

$$R\left(X_{\tau_1^r} \cup X_{\tau_2^r} \cup \ldots \cup X_{\tau_q^r}\right) \cap X_{\tau_j^r} = R\left(X_{\tau_j^r}\right),$$

so $h = \left\{\tau_j^r \vee \vec{1}\right\}$, as required. ■

Lemma 6. *Suppose $Z = \{z_1, z_2, \ldots, z_n\} \subseteq A$ and let $\tau_1^r, \tau_2^r, \ldots, \tau_p^r$ be some maps from a subset of Z of size r to $\{0\}$. Then*

$$R(X_{\tau_1^r} \cup X_{\tau_2^r} \cup \ldots \cup X_{\tau_p^r}) = \left\{\tau_1^r \vee \vec{1}, \tau_2^r \vee \vec{1}, \ldots, \tau_p^r \vee \vec{1}\right\}.$$

Proof. Let $\tau_1^r, \tau_2^r, \ldots, \tau_q^r$ be as in Lemma 5. Then by Obstinacy

$$\begin{aligned} R(X_{\tau_1^r} \cup X_{\tau_2^r} \cup \ldots \cup X_{\tau_p^r}) &= R(X_{\tau_1^r} \cup X_{\tau_2^r} \cup \ldots \cup X_{\tau_q^r}) \cap \left(X_{\tau_1^r} \cup X_{\tau_2^r} \cup \ldots \cup X_{\tau_p^r}\right) \\ &= \left\{\tau_1^r \vee \vec{1}, \tau_2^r \vee \vec{1}, \ldots, \tau_p^r \vee \vec{1}\right\}. \end{aligned}$$

■

We now have all the devices necessary to move on to the crucial step.

Lemma 7. *Let $\tau_1^{r_1}, \tau_2^{r_2}, \ldots, \tau_p^{r_p}$ be maps each from some subset of Z of cardinality $r_1, \ldots r_p$ to $\{0\}$ respectively. If $R(2^A) = \left\{\vec{1}\right\}$, then for $r = \min\{r_i \mid i = 1, \ldots, p\}$*

$$R\left(X_{\tau_1^{r_1}} \cup X_{\tau_2^{r_2}} \cup \ldots \cup X_{\tau_p^{r_p}}\right) = \left\{\tau_j^{r_j} \vee \vec{1} \mid r_j = r\right\}.$$

Proof. Let $\delta_1^r, \delta_2^r, \ldots, \delta_q^r$ be all the maps from a subset of size r of Z to $\{0\}$. Then

$$\left\{\tau_j^{r_j} \vee \vec{1} \mid r_j = r\right\} \subseteq \left\{\delta_i^r \vee \vec{1} \mid i = 1, \ldots, q\right\}. \tag{11}$$

Now, since each $X_{\tau_i^{r_i}} \subseteq X_{\delta_k^r}$, for some k, by lemma 6 above and (11)

$$\begin{aligned} R\left(X_{\tau_1^{r_1}} \cup X_{\tau_2^{r_2}} \cup \ldots \cup X_{\tau_p^{r_p}}\right) &= R(X_{\delta_1^r} \cup X_{\delta_2^r} \cup \ldots \cup X_{\delta_q^r}) \cap \left(X_{\tau_1^{r_1}} \cup X_{\tau_2^{r_2}} \cup \ldots \cup X_{\tau_p^{r_p}}\right) \\ &= \left\{\tau_j^{r_j} \vee \vec{1} \mid r_j = r\right\}. \end{aligned}$$

■

Corollary 8. *For $X \in \mathbb{K}$, if $R(2^A) = \{\vec{1}\}$ then*

$$R(X) = \left\{ f \in X \mid \left| f^{-1}\{0\} \right| = r \right\},$$

where r is minimal such that $|f^{-1}\{0\}| = r$ for some $f \in X$.

Proof. The result follows as an immediate consequence of Obstinacy and Lemma 7. ■

Notice that by duality, Corollary 8 holds for $\vec{1}$ being replaced by $\vec{0}$.

This completes the proof of the "if" direction of Theorem 1. We now move on to show its converse, namely that if a Reason $R(\cdot)$ is defined in any of the above three ways, then Renaming, Irrelevance and Obstinacy are satisfied. This clearly characterizes completely Regulative Reasons for the special case in which worlds are maps from finite set A to 2.

Again, we start with the trivial Reasons, and then we move on to the case of the non-trivial ones.

Lemma 9. *Suppose $R(X) = X$, for all $X \in \mathbb{K}$. Then Renaming, Obstinacy and Irrelevance are satisfied.*

Proof. (Renaming) Suppose $K \in \mathbb{K}$ with support $X \subseteq A$ and π is a permutation of A. Then

$$R(K)\pi = K\pi = R(K\pi)$$

as required.

(Obstinacy) For $K_1, K_2 \in \mathbb{K}$, with supports $X_1, X_2 \subseteq A$ respectively,

$$R(K_1) \cap K_2 = K_1 \cap K_2 = R(K_1 \cap K_2)$$

as required.

(Irrelevance) Suppose $K_1, K_2 \in \mathbb{K}$ (with supports X_1, X_2 respectively) are such that for any $f_1 \in K_1, f_2 \in K_2$, there exists $f_3 \in \mathbb{W}$ such that $f_3 \restriction X_1 = f_1 \restriction X_1$ and $f_3 \restriction X_2 = f_2 \restriction X_2$. We have to show that $R(K_1) \restriction X_1 = R(K_1 \cap K_2) \restriction X_1$. Let $g \in 2^{X_1}$. If $g \in R(K_1 \cap K_2) \restriction X_1$ then obviously $g \in R(K_1) \restriction X_1$. As to the other direction, suppose $g = f_1 \restriction X_1$ with $f_1 \in K_1$. Then we are given that for $f_2 \in K_2$ there is $f_3 \in \mathbb{W}$ such that $f_3 \restriction X_1 = f_1 \restriction X_1 = g$ and $f_3 \restriction X_2 = f_2 \restriction X_2$ Thus, $f_3 \in K_1 \cap K_2$ and $g = f_3 \restriction X_1 \in R(K_1 \cap K_2) \restriction X_1$, as required. ■

Lemma 10. *$R_1(K)$ satisfies Renaming, Obstinacy and Irrelevance.*

Proof.
(Renaming) Let σ be a permutation of A. Then

$$\begin{aligned} f \in R_1(K)\sigma &\iff f = g\sigma, \text{for some } g \in R_1(K) \\ &\iff f = g\sigma, \text{for some } g \in \left\{h \in K \mid \left|h^{-1}\{1\}\right| = r\right\} \end{aligned} \tag{12}$$

where $r = \max\{|h^{-1}\{1\}| \mid h \in K\}$. But since $|h^{-1}\{1\}| = |(h\sigma)^{-1}\{1\}|$, then

$$h \in K \text{ and } \left|h^{-1}\{1\}\right| = r \iff h\sigma \in K\sigma \text{ and } \left|(h\sigma)^{-1}\{1\}\right| = r.$$

and $r = \max\{|(h\sigma)^{-1}\{1\}| \mid h\sigma \in K\sigma\}$. Hence

$$f \in R_1(X) \iff f\sigma \in R_1(X\sigma),$$

as required.
(Obstinacy) Let $K_1, K_2 \in \mathbb{K}$ and let $R_1(K_1) \cap K_2 \neq \emptyset$ and set

$$r' = \max\left\{\left|g^{-1}\{1\}\right| \mid g \in K_1 \cap K_2\right\}$$

.

We claim that $r' = r$, where r is defined as above. To see that the result follows from this claim notice that if $r' = r$, then

$$\begin{aligned} R(K_1 \cap K_2) &= \left\{f \in K_1 \cap K_2 \mid \left|f^{-1}\{1\}\right| = r\right\} \\ &= \left\{f \in K_1 \mid \left|f^{-1}\{1\}\right| = r\right\} \cap K_2 \\ &= R_1(K_1) \cap K_2. \end{aligned}$$

We show the claim by contradiction. Since $K_1 \cap K_2 \subseteq K_1$, the case $r' > r$ is clearly not possible. To see that $r' < r$ is not possible either, and hence that $r' = r$, let $h \in R_1(K_1) \cap K_2$. Then r' would be the largest n for which there exists $h' \in K_1 \cap K_2$ such that $\left|h'^{-1}\{1\}\right| = n$. But since $h \in R_1(K_1)$, r would be such an n, giving $r' \geq r$ as required.

(Irrelevance) Suppose $K_1, K_2 \in \mathbb{K}$ (with supports X_1, X_2, respectively) and for any $f_1 \in K_1, f_2 \in K_2$, there exists $f_3 \in \mathbb{W}$ such that $f_3 \restriction X_1 = f_1 \restriction X_1$ and $f_3 \restriction X_2 = f_2 \restriction X_2$. We have to show that

$$R_1(K_1) \restriction X_1 = R_1(K_1 \cap K_2) \restriction X_1.$$

So assume that $g \in R_1(K_1) \restriction X_1$. Then $\exists f_1 \in R_1(K_1)$ such that $f_1 \restriction X_1 = g$. We now claim that

$$\forall x \notin X_1 \quad f_1(x) = 1. \tag{13}$$

Suppose otherwise and define

$$f'(x) = \begin{cases} f_1(x) & \text{if} \quad x \in X_1 \\ \\ 1 & \text{otherwise.} \end{cases}$$

Then $f' \in K_1$ but $\left|f'^{-1}\{1\}\right| > \left|f_1^{-1}\{1\}\right|$, which is impossible if $f_1 \in R_1(K_1)$. Hence $X_1 \supseteq \{x \mid f_1(x) = 0\}$ (and similarly, $X_2 \supseteq \{x \mid f_2(x) = 0\}$, for $f_2 \in R_1(K_2)$). Thus $\exists f \in K_1 \cap K_2$ such that $f \restriction X_1 = f_1 \restriction X_1$ and $f \restriction X_2 = f_2 \restriction X_2$. Moreover, since $X_1 \cup X_2$ is a support for $K_1 \cap K_2$, can also assume that

$$f(x) = 1, \quad \text{for all} \quad x \notin X_1 \cup X_2. \tag{14}$$

Claim now that there is no $h \in K_1 \cap K_2$ such that

$$\left|h^{-1}\{1\}\right| > \left|f^{-1}\{1\}\right|. \tag{15}$$

Suppose on the contrary that such an h existed. By (14) we may assume $h(x) = 1$ for all $x \notin X_1 \cup X_2$. Notice first that

$$x \in X_1 \cap X_2 \Rightarrow f(x) = h(x). \tag{16}$$

To see this, notice that $f \in K_1, h \in K_2$. So $\exists g'$ such that $g' \restriction X_1 = f \restriction X_1$ and $g' \restriction X_2 = h \restriction X_2$. Hence $f(x) = g'(x) = h(x)$, as required. Now,

$$\begin{aligned} \left|h^{-1}\{1\}\right| = \overbrace{\left|\{y \in X_1 - X_2 \mid h(y) = 1\}\right|}^{\alpha^h} + \left|\{y \in X_2 - X_1 \mid h(y) = 1\}\right| + \\ + \left|\{y \in X_2 \cap X_1 \mid h(y) = 1\}\right|. \end{aligned}$$

and

$$\begin{aligned} \left|f^{-1}\{1\}\right| = \overbrace{\left|\{y \in X_1 - X_2 \mid f(y) = 1\}\right|}^{\alpha^f} + \left|\{y \in X_2 - X_1 \mid f(y) = 1\}\right| + \\ + \left|\{y \in X_2 \cap X_1 \mid f(y) = 1\}\right|. \end{aligned}$$

Without loss of generality then, if $|h^{-1}\{1\}| > |f^{-1}\{1\}|$ then $\alpha^h > \alpha^f$. But this leads to the required contradiction. To see that define

$$h'(z) = \begin{cases} h(z) & \text{if} \quad z \in X_1 \\ \\ 1 & \text{otherwise.} \end{cases}$$

Then $h' \in K_1$ but $\left|h'^{-1}\{1\}\right| = |h^{-1}\{1\} \cap X_1| > \left|f_1^{-1}\{1\}\right|$, and this is clearly inconsistent with $f_1 \in R(K_1)$. So $f \in R(K_1 \cap K_2)$ and hence $g \in R(K_1 \cap K_2)\restriction X_1$, as required for this direction of the proof.

As to the other direction for Irrelevance , assume that $g \in R(K_1 \cap K_2)\restriction X_1$ but $g \notin R(K_1)\restriction X_1$. Define

$$g'(x) = \begin{cases} g(x) & \text{if} \quad x \in X_1 \\ \\ 1 & \text{otherwise.} \end{cases}$$

Then, $g' \in K_1$ as it agrees on X_1 with $g \in K_1$. Indeed $g' \notin R(K_1)\restriction X_1$ too, since $g'\restriction X_1 = g\restriction X_1$. Hence $\exists f \in R(K_1)$ such that

$$|\{y \in X_1 \mid f(y) = 1\}| > |\{y \in X_1 \mid g(y) = 1\}|\,. \tag{17}$$

Now pick $h \in R(K_1 \cap K_2)$ such that $h \restriction X_1 = g$ and define f' such that $f'\restriction X_1 = f\restriction X_1$ and $f'\restriction X_2 = h\restriction X_2$. As above we can assume that

$$f'(x) = 1 \,\text{for all}\, x \notin X_1 \cup X_2 \tag{18}$$

Then $f' \in K_1 \cap K_2$ and $\left|f'^{-1}\{1\} \cap X_1\right| > |h^{-1}\{1\} \cap X_1|$ (by (17) and the facts $f' \restriction X_1 = f \restriction X_1$ and $h \restriction X_1 = g$). Thus, since $\left|f'^{-1}\{1\} \cap X_2\right| = |h^{-1}\{1\} \cap X_2|$ and $f'\restriction X_1 \cap X_2 = h\restriction X_1 \cap X_2$, we have that

$$\left|f'^{-1}\{1\} \cap (X_1 \cup X_2)\right| > \left|h^{-1}\restriction\{1\} \cap (X_1 \cup X_2)\right|.$$

But this is inconsistent with the maximality of $|h^{-1}\{1\}|$, concluding the proof of the converse of Theorem 1.

■

A pleasing aspect of Theorem 1 is that it seems to us to point to precisely the answer(s) that people commonly do come up with when presented with this choice problem. For example in the case

$$\begin{array}{cccc} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{array}$$

it is our experience that the fourth row, 1 1 1 1, is the favored choice. In other words the (unique) choice according to R_1. Of course that is not the only Regulative Reason, R_0 gives $\{ 0001, 0100, 0010 \}$ whilst the trivial reason gives us back the whole set. Clearly though those two Reasons could be seen as inferior to R_1 here because they ultimately require a random choice from a larger set, thus increasing the probability of non-agreement. (This idea will be explored further in the next chapter when we come to Reasons based on Ambiguity.) This seems to point to a further elaboration of our picture whereby the agent might for a particular K experiment with several Reasons and ultimately settle for a choice which depends on K itself[5]. We shall return to this point later.

Of course one might argue in this example that in making the choice of 1 1 1 1 one was not *consciously* aware of any obligation to satisfy Renaming, Obstinacy and Irrelevance. Be that as it may it is nevertheless interesting we feel that observance of these principles turns out to be both so restrictive and to rather frequently leads to 'the people's choice'. Notice too that if one does adopt a Regulative Reason then one automatically also observes Obstinacy. This *could* then be offered as a defense of Obstinacy against the earlier criticism, that it is no more unreasonable than adopting a Regulative Reason. Whether or not there are alternative sets of 'justified' principles which yield interesting families of reasons such as the one we have considered here remains a matter for further investigation.

4 The Minimum Ambiguity Reason

4.1 An Informal Procedure

In the previous section we saw how an agent might arrive at a particular canonical Reason by adopting and adhering to certain principles, principles which (after some consideration) one might suppose any other like-minded agent might similarly come to. An alternative approach, which we shall investigate in this section, is to introduce a notion of 'distinguishability', or 'indistinguishability', between elements of K and chose as $R(K)$ those most distinguished, equivalently *least ambiguous*, elements. Instead of being based on principles this $R(K)$ will in the first instance be specified by a procedure, or algorithm, for constructing it.

[5]Alternatively one might hedge one's bets and adopt the "collected extremal choice", $R_\cup(K) = R_0(K) \cup R_1(K)$, in the sense of [1] (see also [25], p. 163) and by the Aizerman-Malishevski Theorem (Theorem 4 of [1]) $R_\cup$ is a *Plott function*, that is to say a function that satisfies the so-called Path Independence property introduced in [23].

The idea behind the construction of $R(K)$ is based on trying to fulfill two requirements. The first requirement is that if f and g are, as elements of K, *indistinguishable*, then $R(K)$ should not contain one of them, f say, without also containing the other, g. In other words an agent should not give positive probability to picking one of them but zero probability to picking the other. The argument for this is that if they are 'indistinguishable' on the basis of K then another agent could just as well be making a choice of $R(K)$ which included g but not f. Since agents are trying to make the same ultimate choice of element of K this surely looks like an undesirable situation (and indeed, as will later become clear, taking that route may be worse, and will never be better, than avoiding it).

According to this first requirement then $R(K)$ should be closed under the 'undistinguishability relation'.

The second requirement is that the agent's choice of $R(K)$ should be as small as possible (in order to maximize the probability of randomly picking the same element as another agent) subject to the additional restriction that this way of thinking should not equally permit another like-minded agent (so also, globally, satisfying the first requirement) to make a different choice, since in that case any advantage of picking from the small set is lost.

The first consequence of this is that initially the agent should be looking to choose from those minimal subsets of K closed under indistinguishability, 'minimal' here in the sense that they do not have any proper non-empty subset closed under indistinguishability. Clearly if this set has a unique smallest element then the elements of this set are the least ambiguous, most outstanding, in K and this would be a natural choice for $R(K)$. However, if there are two or more potential choices $X_1, X_2, ..., X_k$ at this stage with the same number of elements then the agent could do no worse than combine them into a single potential choice $X_1 \cup X_2 \cup ... \cup X_k$ since the choice of any one of them would be open to the obvious criticism that another 'like-minded agent' could make a different (in this case disjoint) choice, which would not improve the chances of a match (and may make them considerably worse if the first agent subsequently rejected $X_1 \cup X_2 \cup ... \cup X_k$ in favor of a better choice). Faced with this revelation our agent would realize that the 'smallest' way open to reconcile these alternatives is to now permit $X_1 \cup X_2 \cup ... \cup X_k$ as a potential choice whilst dropping $X_1, X_2, ..., X_k$[6].

The agent now looks again for a smallest element from the current set of potential choices and carries on arguing and introspecting in this way until

[6]It is noted in [25] that this strategy mirrors the "sceptical" as opposed to the "credulous" approach to non-monotonic inference.

eventually at some stage a unique choice presents itself.

In what follows we shall give a formalization of this procedure.

4.2 Permutations and ambiguity

The first step in the construction of the Minimum Ambiguity Reason consists in providing the agent with a notion of equivalence or indistinguishability among worlds in a given $K \subseteq 2^A$.

In fact with the minimal structure we have available here the notion we want is almost immediate: Elements g, h of K are *indistinguishable* (with respect to K) if there is a permutation σ of A such that

$$K = K\sigma \ (= \{ f\sigma \mid f \in K \})$$

and $g\sigma = h$.

We shall say that a permutation σ of A is a permutation of K if $K = K\sigma$.

The idea here is that within the context of our choice problem a permutation σ of K maps $f \in K$ to an $f\sigma$ in $K\sigma$ which has essentially the standing within $K\sigma$ $(= K)$ as f had within K. In other words as far as K is concerned f and $f\sigma$ are indistinguishable. The following Lemma is immediate.

Lemma 11. *If σ and τ are permutations of K then so are $\sigma\tau$ and σ^{-1}.*

Having now disposed of what we mean by indistiguishability between elements of $K \subseteq 2^A$, we now recursively define for $f \in K$ *the ambiguity class of f within K at level m* by:

$$\mathbb{S}_0(K, f) = \{g \in K \mid \exists \text{ permutation } \sigma \text{ of } K \text{ such that } f\sigma = g\}$$

$$\mathbb{S}_{m+1}(K, f) = \begin{cases} \{g \in K \mid |\mathbb{S}_m(K, f)| = |\mathbb{S}_m(K, g)|\} & \text{if } |\mathbb{S}_m(K, f)| \leq m + 1, \\ \mathbb{S}_m(K, f) & \text{otherwise.} \end{cases}$$

For $f, g \in K$ define the relation

$$g \sim_m f \Leftrightarrow g \in \mathbb{S}_m(K, f).$$

Lemma 12. *$\sim_m$ is an equivalence relation.*

Proof. By induction on m. For the case $m = 0$ this is clear since if $f, g, h \in K$ and $f\sigma = g$, $g\tau = h$ with σ, τ permutations of K then $g\sigma^{-1} = f$, $f\sigma\tau = h$ and by Lemma 11 $\sigma^{-1}, \sigma\tau$ are also permutations of K.

Assume true for m. If $|\mathbb{S}_m(K,f)| > m+1$ then, by the definition of $\mathbb{S}_{m+1}(K,f)$, the result follows immediately from the inductive hypothesis. Otherwise, the reflexivity of $\sim_m$ is again immediate. For symmetry assume that $g \in \mathbb{S}_{m+1}(K,f)$. Then $g \in \{h \in K \mid |\mathbb{S}_m(K,h)| = |\mathbb{S}_m(K,f)|\}$, so $|\mathbb{S}_m(K,g)| = |\mathbb{S}_m(K,f)|$ and $f \in \{h \in K \mid |\mathbb{S}_m(K,h)| = |\mathbb{S}_m(K,g)|\}$. An analogous argument shows that $\sim_{m+1}$ is also transitive. ■

Thus, as f ranges over K, $\sim_m$ induces a partition on K and the sets $\mathbb{S}_m(K,f)$ are its equivalence classes. Moreover, this m-th partition is a refinement of the $m+1$-st partition. In other words, the sets $\mathbb{S}_m(K,f)$ are increasing and so eventually constant fixed at some set which we shall call $\mathbb{S}(K,f)$.

The *ambiguity of f within K* is then defined by:

$$\mathbb{A}(K,f) =_{def} |\mathbb{S}(K,f)|.$$

Finally, we can define the *Minimum Ambiguity Reason* $R_{\mathbb{A}}(K)$ by letting:

$$R_{\mathbb{A}}(K) = \{f \in K \mid \forall g \in K,\ \mathbb{A}(K,f) \leq \mathbb{A}(K,g)\}. \tag{19}$$

As a rather self evident consequence of the definition of $R_{\mathbb{A}}$ we have the following result.

Proposition 13. *$R_{\mathbb{A}}(K) = \mathbb{S}(K,f)$, for any $f \in R_{\mathbb{A}}(K)$*

Proof. Let $f \in R_{\mathbb{A}}(K)$. To show that $\mathbb{S}(K,f) \subseteq R_{\mathbb{A}}(K)$ suppose $\mathbb{S}(K,f) = \mathbb{S}_m(K,f)$ and $g \in \mathbb{S}_m(K,f)$. Then $\mathbb{S}_m(K,g) = \mathbb{S}_m(K,f)$ so $\mathbb{S}_m(K,g)$ must equal $\mathbb{S}(K,g)$ (since m could be taken arbitrarily large) and $|\mathbb{S}(K,g)| = |\mathbb{S}(K,f)|$, so $g \in R_{\mathbb{A}}(K)$. Conversely let $g \in R_{\mathbb{A}}(K)$ and fix some large m. If $g \notin \mathbb{S}(K,f)$, then $\mathbb{S}(K,f) \cap \mathbb{S}(K,g) = \emptyset$ and since both f and g are in $R_{\mathbb{A}}(K)$, then $|\mathbb{S}(K,f)| = |\mathbb{S}(K,g)|$. But this leads to the required contradiction since for m large enough, $|\mathbb{S}_m(K,f)| \leq m+1$ so $\mathbb{S}_m(K,f)$ and $\mathbb{S}_m(K,g)$ would both be proper subsets of $\mathbb{S}_{m+1}(K,f)$. Thus g would eventually be in $\mathbb{S}_m(K,f)$, contradicting the hypothesis. ■

Example Let $K \in \mathbb{K}$ and suppose that as f ranges over K the 0-ambiguity classes of f in K are given by the following partition of K

$\{a_1, a_2\}, \{b_1, b_2\}, \{c_1, c_2\},$
$\{d_1, d_2, d_3\}, \{e_1, e_2, e_3\},$
$\{f_1, f_2, ..., f_6\}, \{g_1, g_2, ..., g_6\},$
$\{h_1, h_2, ..., h_{12}\},$
$\{i_1, i_2, ..., i_{24}\}.$

For $m = 1$ the classes remain fixed. For $m = 2$ the first three classes get combined and the $\mathbb{S}_2(K, f)$ look like

$\{a_1, a_2, b_1, b_2, c_1, c_2\}$,
$\{d_1, d_2, d_3\}, \{e_1, e_2, e_3\}$,
$\{f_1, f_2, ..., f_6\}, \{g_1, g_2, ..., g_6\}$,
$\{h_1, h_2, ..., h_{12}\}$,
$\{i_1, i_2, ..., i_{24}\}$.

Similarly for $m = 3$ where the two classes of size 3 are combined so that the $\mathbb{S}_3(K, f)$ become

$\{a_1, a_2, b_1, b_2, c_1, c_2\}$,
$\{d_1, d_2, d_3, e_1, e_2, e_3\}$,
$\{f_1, f_2, ..., f_6\}, \{g_1, g_2, ..., g_6\}$,
$\{h_1, h_2, ..., h_{12}\}$,
$\{i_1, i_2, ..., i_{24}\}$.

The ambiguity classes do not change until step 6 when the four classes with 6 elements are combined making $\mathbb{S}_6(K, f)$ look like

$\{a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2, d_3, e_1, e_2, e_3, f_1, f_2, ..., f_6, g_1, g_2, ..., g_6\}$,
$\{h_1, h_2, ..., h_{12}\}$,
$\{i_1, i_2, ..., i_{24}\}$.

Finally, we combine the two classes with 24 elements and obtain $\mathbb{S}_{24}(K, f)$ with just two classes

$\{a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2, d_3, e_1, e_2, e_3, f_1, f_2, ..., f_6, g_1, g_2, ..., g_6, i_1, i_2, ..., i_{24}\}$,
$\{h_1, h_2, ..., h_{12}\}$.

Clearly the ambiguity classes stabilize at this 24-th step and hence the Minimum Ambiguity Reason for this K gives the 12-set $\{h_1, h_2, ..., h_{12}\}$.

Notice that, in the definition of the ambiguity classes of K, the splitting of the inductive step into two cases is indeed necessary to ensure that some sets closed under permutations of K are not dismissed unnecessarily early. This same example shows that if we allowed the inductive step in the definition to be replaced by the (somehow more intuitive) equation

$$\mathbb{S}_{m+1}(K, f) = \{g \in K \mid |\mathbb{S}_m(K, f)| = |\mathbb{S}_m(K, g)|\} \tag{20}$$

we would fail to pick the "obvious" smallest such subset of K. To see this suppose again that K is as above but this time the alternative procedure based on (20) was used to construct $R_{\mathbb{A}}$. Then we would have all the classes of the same size all merged in one step so that the $1-$ambiguity classes $\mathbb{S}_1(K, f)$ would look like:

$\{a_1, a_2, b_1, b_2, c_1, c_2\}$,
$\{d_1, d_2, d_3, e_1, e_2, e_3\}$,
$\{f_1, f_2, ..., f_6, g_1, g_2, ..., g_6\}$,
$\{h_1, h_2, ..., h_{12}\}$,
$\{i_1, i_2, ..., i_{24}\}$.

Then $\mathbb{S}_2(K, f)$ would look like this:

$\{a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2, d_3, e_1, e_2, e_3\}$,
$\{f_1, f_2, ..., f_6, g_1, g_2, ..., g_6, h_1, h_2, ..., h_{12}\}$,
$\{i_1, i_2, ..., i_{24}\}$,

so that the procedure stabilizes at $m = 3$ with $\mathbb{S}(K, f)$ of the form:

$\{a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2, d_3, e_1, e_2, e_3\}$,
$\{f_1, f_2, ..., f_6, g_1, g_2, ..., g_6, h_1, h_2, ..., h_{12}, i_1, i_2, ..., i_{24}\}$,

Hence, the construction that follows the alternative definition of ambiguity classes, which imposes no restriction on appropriate stage for the combination of the classes, leads again toa 12-set. However, this alternative procedure appears to miss out what naturally seems to be a more distinguished subset of K.

4.3 Justifying the Minimum Ambiguity Reason

We now want to show that the Minimum Ambiguity Reason defined in (19) is an adequate formalization of the informal description given in section 4.1. Recall that we put forward two informal desiderata for the resulting selection from K, firstly that it should be closed under indistinguishability and secondly that it should be the unique smallest possible such subset not eliminated by there being a like-minded agent who by similar reasoning could arrive at a different answer.

As far as the former is concerned notice that by Lemma 13 $R_{\mathbb{A}}(K)$ is closed under all the $\sim_m$, not just $\sim_0$. Thus this requirement of closure under indistinguishability is met, *assuming* of course that one accepts this interpretation of 'indistinguishability'. Indeed $R_{\mathbb{A}}$ satisfies Renaming as we now show.

Theorem 14. *$R_{\mathbb{A}}$ satisfies Renaming.*

Proof. As usual let σ be a permutation of A. We need to prove that

$$R_{\mathbb{A}}(K)\sigma = R_{\mathbb{A}}(K\sigma).$$

We first show by induction on m that for all $f \in K$, $\mathbb{S}_m(K, f)\sigma = \mathbb{S}_m(K\sigma, f\sigma)$. To show the base case $m = 0$ for all $f \in K$, let

$$\mathbb{S}_0(K, f) = \{\, g_1, \ldots, g_q \,\}.$$

Choose a permutation τ of K such that $f\tau = g_i$. Then $\sigma^{-1}\tau\sigma$ is a permutation of $K\sigma$ and $(f\sigma)\sigma^{-1}\tau\sigma = g_i\sigma$. Hence, $\mathbb{S}_0(K,f)\sigma \subseteq \mathbb{S}_0(K\sigma, f\sigma)$. Similarly, $\mathbb{S}_0(K\sigma, f\sigma)\sigma^{-1} \subseteq \mathbb{S}_0(K,f)$, so equality must hold here.

Assume now the result for the $\mathbb{S}_m$-th ambiguity class, so we want to prove that

$$\mathbb{S}_{m+1}(K,f)\sigma = \mathbb{S}_{m+1}(K\sigma, f\sigma).$$

We distinguish between two cases, corresponding to the ones appearing in the construction of the ambiguity classes. Recall that $\mathbb{S}_{m+1}(K,f) = \mathbb{S}_m(K,f)$ if $m+1 > |\mathbb{S}_m(K,f)|$. So, in this case, the result follows immediately by the inductive hypothesis. Otherwise, since σ (on 2^A) is 1-1, it is enough to see that

$$\begin{aligned}
\mathbb{S}_{m+1}(K,f)\sigma &= \{g \in K \mid |\mathbb{S}_m(K,f)| = |\mathbb{S}_m(K,g)|\}\sigma \\
&= \{g\sigma \in K\sigma \mid |\mathbb{S}_m(K\sigma, f\sigma)| = |\mathbb{S}_m(K\sigma, f\sigma)|\} \text{ (i.h.)} \\
&= \mathbb{S}_{m+1}(K\sigma, f\sigma).
\end{aligned}$$

Since, by Lemma 13, $R_{\mathbb{A}}(K)$ is the smallest $\mathbb{S}(K,f)$, this concludes the proof of the Lemma. ■

Before further considering how far our formal construction of $R_{\mathbb{A}}(K)$ matches the informal description in section 3.1, it will be useful to have the next result to hand.

Theorem 15. *A non-empty $K' \subseteq K$ is closed under permutations of K into itself if and only if there exists a Reason R satisfying Renaming such that $R(K) = K'$.*

Proof. The direction from right to left follows immediately from the Renaming principle. For the other direction define, for $K_1 \subseteq 2^A, K_1 \neq \emptyset$,

$$R(K_1) = \begin{cases} K'\sigma & \text{if } K_1 = K\sigma \text{ for some permutation } \sigma \text{ of } A; \\ K_1 & \text{otherwise .} \end{cases} \tag{21}$$

Note that in the first case $R(K_1)$ is defined unambiguously, that is to say, whenever we have two permutations σ_1, σ_2 of A such that $K_1 = K\sigma_1 = K\sigma_2$, then $K'\sigma_1 = K'\sigma_2$. This follows since in this case, $\sigma_2\sigma_1^{-1}$ is a permutation of A and $K\sigma_2\sigma_1^{-1} = K$ so $K'\sigma_2\sigma_1^{-1} = K'$, i.e. $K'\sigma_1 = K'\sigma_2$.

We now want to show that if σ is a permutation of A and $K_1\sigma = K_2$ then $R(K_2) = R(K_1)\sigma$. If K_1 is covered by the first case of (21), then so is K_2, for if τ is a permutation of A such that $K_1 = K\tau$, then $K_2 = K\tau\sigma$ and $R(K_1\sigma) = R(K_2) = K'\tau\sigma = R(K_1)\sigma$. If K_1 is covered by the second case of (21), so is K_2 since if $K_2 = K\tau$ for some permutation τ of A, then

$K_1 = K\tau\sigma^{-1}$ so $R(K_1)$ would be defined by the first case. It follows then that here we must have $R(K_1\sigma) = R(K_2) = K_2 = K_1\sigma = R(K_1)\sigma$ as required. ■

The importance of this result is that in the construction of $R_{\mathbb{A}}(K)$ the choices $\mathbb{S}_m(K, f)$ which were eliminated (by coalescing) because of there currently being available an alternative choice of a $\mathbb{S}_m(K, g)$ of the same size are indeed equivalently being eliminated on the grounds that there is a like-minded agent, even one satisfying Renaming, who could pick $\mathbb{S}_m(K, g)$ in place of $\mathbb{S}_m(K, f)$. In other words it is not as if some of these choices are barred because no agent could make them whilst still satisfying Renaming. Once a level m is reached at which there is a unique smallest $S_m(K, f)$ this will be the choice for the informal procedure. It is also easy to see that this set will remain the unique smallest set amongst all the subsequent $S_n(K, g)$, and hence will qualify as $R_{\mathbb{A}}(K)$. In this sense then our formal procedure fulfills the intentions of the informal description of section 3.1.

4.4 Comparing Regulative and Minimum Ambiguity Reasons

In this and the previous section we have put forward arguments for both the Regulative and Minimum Ambiguity Reasons being considered as 'rational' within the understanding of that term in this paper. Interestingly in practice neither seems to come out self evidently better in all cases. For example, in the case considered earlier of

$$\begin{array}{cccc} 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{array}$$

R_1 gives the singleton $\{1111\}$ whilst $R_{\mathbb{A}}$ gives the somewhat unexceptional $\{0011, 1100\}$ and R_0 the rather useless $\{0011, 0110, 1100\}$. On the other hand if we take the subset

$$\begin{array}{cccc} 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{array}$$

of this set $R_{\mathbb{A}}$ gives $\{0110\}$ whilst both R_1 and R_0 give the whole set.

Concerning the defining principles of the Regulative Reasons, whilst as we have seen $R_{\mathbb{A}}$ does satisfy Renaming the above example shows that it fails

to satisfy Obstinacy. Indeed with a little more work we can show that it does not even satisfy Idempotence, that is $R(R(K)) = R(K)$, a consequence of Obstinacy. Finally $R_{\mathbb{A}}$ does not satisfy Irrelevance either. For an example to show this let K_1 consist of

$$\begin{array}{cccccccccccc}
1 & 0 & 0 & 1 & * & * & * & * & * & * & * \\
1 & 1 & 0 & 1 & * & * & * & * & * & * & * \\
1 & 1 & 1 & 1 & * & * & * & * & * & * & * \\
1 & 0 & 0 & 0 & * & * & * & * & * & * & * \\
0 & 0 & 0 & 1 & * & * & * & * & * & * & *
\end{array}$$

and let K_2 consist of

$$\begin{array}{cccccccccccc}
* & * & * & * & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
* & * & * & * & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
* & * & * & * & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
* & * & * & * & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
* & * & * & * & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
* & * & * & * & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
* & * & * & * & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
* & * & * & * & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
* & * & * & * & 0 & 0 & 0 & 0 & 0 & 1 & 1
\end{array}$$

where $*$ indicates a free choice of 0 or 1. Then K_1, K_2 satisfy the requirements of Irrelevance and $R_{\mathbb{A}}(K_1), R_{\mathbb{A}}(K_2)$ are respectively

$$\begin{array}{cccccccccccc}
 & & & & & & & & & & & \\
1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}
\qquad
\begin{array}{cccccccccccc}
1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0
\end{array}$$

whereas $R_{\mathbb{A}}(K_1 \cap K_2)$ is

$$\begin{array}{cccccccccccc}
1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0
\end{array}$$

5 The Smallest Uniquely Definable Reason

In this section we present another Reason which, at first sight, looks a serious challenger to the Regulative and Minimum Ambiguity Reasons so far introduced.

Consider again an agent who is given a non-empty subset K of 2^A from which to attempt to make a choice which is common to another like-minded agent. A natural approach here might be for the agent to consider all non-empty subsets of K that could be described, or to use a more formal term, defined, within the *structure available* to the agent. If some individual element was definable (meaning definable in this structure *without parameters*) then this would surely be a natural choice, unless of course there were other such elements. Similarly choosing a small definable set and then choosing randomly from within it would seem a good strategy, *provided there were no other definable sets of the same size.* Reasoning along these lines then suggests that our agent could reach the conclusion that s/he should choose the smallest definable set for which there was no other definable set of the same size.

Of course all this depends on what we take to be the *structure available* to the agent. In what follows we shall consider the case when the agent can recognize 0 and 1, elements of A, $\{0,1\}$ and K, composition and equality[7]. Precisely, let $\mathcal{M}$ be the structure

$$\langle \{0,1\} \cup A \cup K, \{0,1\}, A, K, =, Comp, 0, 1 \rangle$$

where $=$ is equality for $\{0,1\} \cup A \cup K$ (we assume of course that A, $\{0,1\}$, 2^A are all disjoint) and $Comp$ is a binary function which on $f \in K$, $a \in A$ gives $f(a)$ (and, say, the first coordinate on arguments not of this form). As usual we shall write $f(a) = i$ in place of $Comp(f, a) = i$ etc..

We define the Uniquely Smallest Definable Reason, $R_{\mathbb{U}}$, by setting $R_{\mathbb{U}}(K)$ to be that smallest $\emptyset \neq K' \subseteq K$ first order definable in $\mathcal{M}$ for which there is no other definable subset of the same size.

The results that follow are directed towards understanding the structure of $R_{\mathbb{U}}(K)$ and its relationship to $R_{\mathbb{A}}(K)$.

[7] One might subsequently argue that the agent could then also recognize automorphisms of $\mathcal{M}$ so the set of these too should be added to our structure, and the whole process repeated, and repeated ... In fact this does not change the definable subsets of K so it turns out there is no point in going down this path.

Lemma 16. *Every permutation σ_0 of K determines an automorphism j_{σ_0} of $\mathcal{M}$ given by the identity on $\{0,1\}$ and*

$$a \in A \longmapsto \sigma_0^{-1}(a), \tag{22}$$

and

$$f \in K \longmapsto f\sigma_0. \tag{23}$$

Conversely every automorphism j_0 of $\mathcal{M}$ determines a permutation σ_{j_0} of K given by

$$\sigma_{j_0}(a) = j_0^{-1}(a) \tag{24}$$

for $a \in A$.

Furthermore for $f \in K$, $f\sigma_{j_0} = j_0(f)$ and the corresponding automorphism determined by $j_{\sigma_{j_0}}$ is j_0 again.

Proof. For σ_0 a permutation of K it is clear that j_{σ_0} defined by (22) and (23) gives a 1-1 onto mapping from A and K into themselves. All that remains to show this first part is to notice that by direct substitution,

$$j_{\sigma_0}(Comp(f,a)) = Comp(f,a) = f(a) = f\sigma_0(\sigma_0^{-1}(a)) = Comp(j_{\sigma_0}(f), j_{\sigma_0}(a)).$$

In the other direction let j_0 be an automorphism of $\mathcal{M}$ and define σ_{j_0} by (24). Then since j_0 is an automorphism of $\mathcal{M}$, σ_{j_0} is a permutation of A and for $f \in K$, $a \in A$,

$$f(a) = j_0(f(a)) = j_0(Comp(f,a)) = Comp(j_0(f), j_0(a)),$$

equivalently,

$$f(a) = j_0(f)(j_0(a)) = j_0(f\sigma_{j_0}^{-1})(a).$$

Hence

$$j_0^{-1}(f)(a) = f\sigma_{j_0}^{-1}(a)$$

so $\sigma_{j_0}^{-1}$ (and hence σ_{j_0} by Lemma 11) is a permutation of K since $j_0^{-1}(f) \in K$, as required.

The last part now follows immediately from the definitions (22), (23), (24). ■

We say that $K' \subseteq K$ satisfies Renaming within K if for all permutations σ of K, $K' = K'\sigma$. Thus 'standard Renaming' is just Renaming within 2^A.

Theorem 17. *A non-empty subset K' of K is definable (without parameters) in $\mathcal{M}$ if and only if K' satisfies Renaming within K.*

Proof. Suppose that K' is definable in $\mathcal{M}$. Then clearly K' is fixed under all automorphisms of $\mathcal{M}$. In particular if σ is a permutation of K then by Lemma 16 j_σ is an automorphism of $\mathcal{M}$ so

$$K' = j_\sigma(K') = K'\sigma$$

Conversely suppose that K' satisfies Renaming within K. Then since every automorphism of $\mathcal{M}$ is of the form j_σ for some permutation σ of K and $j_\sigma(K') = K'\sigma = K'$ it follows that K' is fixed under all automorphisms of $\mathcal{M}$. Consider now the types $\theta^i_1(x), \theta^i_2(x), \theta^i_3(x), ...$ of the elements f_i of K in $\mathcal{M}$. If there were $f_i \in K'$ and $f_j \notin K'$ with the same type then by a back and forth argument (see for example [16]) we could construct an automorphism of $\mathcal{M}$ sending f_i to f_j, contradicting the fact that K' is fixed under automorphisms. It follows that for some n the formulae $\theta^i_1(x) \wedge \theta^i_2(x) \wedge ... \wedge \theta^i_n(x)$ and $\theta^j_1(x) \wedge \theta^j_2(x) \wedge ... \wedge \theta^j_n(x)$ are mutually contradictory when $f_i \in K'$ and $f_j \notin K'$. From this it clearly follows that the formula

$$\bigvee_{f_i \in K'} \bigwedge_{m=1}^{n} \theta^i_m(x)$$

defines K' in $\mathcal{M}$ for suitably large n. ■

Corollary 18. *The sets $\mathbb{S}_m(K, f)$ are definable in $\mathcal{M}$*

Proof. These sets are clearly closed under permutations of K so the result follows from Theorem 17. ■

Theorem 19. *For all $K \in \mathbb{K}$, $|R_{\mathbb{A}}(K)| \leq |R_{\mathbb{U}}(K)|$, with equality just if $R_{\mathbb{A}}(K) = R_{\mathbb{U}}(K)$.*

Proof. We shall show that for all m. If $f \in R_{\mathbb{U}}(K)$ then $\mathbb{S}_m(K, f) \subseteq R_{\mathbb{U}}(K)$. For $m = 0$ this is clear since $R_{\mathbb{U}}(K)$, being definable must be closed under permutations of K. Assume the result for m and let $f \in R_{\mathbb{U}}(K)$. If $\mathbb{S}_{m+1}(K, f)$ were not a subset of $R_{\mathbb{U}}(K)$ there would be $g \in K$ such that $|\mathbb{S}_m(K, f)| = |\mathbb{S}_m(K, g)|$ but $g \notin R_{\mathbb{U}}(K)$. Indeed $\mathbb{S}_m(K, g)$ would have to be entirely disjoint from $R_{\mathbb{U}}(K)$ by the inductive hypothesis. By Corollary 18 $\mathbb{S}_m(K, f)$ and $\mathbb{S}_m(K, g)$ are both definable, and hence so is

$$R_{\mathbb{U}}(K) \cup \mathbb{S}_m(K, g) - \mathbb{S}_m(K, f).$$

But this set is different from $R_{\mathbb{U}}(K)$ yet has the same size, contradiction.

Having established this fact we notice that for $f \in R_{\mathbb{U}}(K)$ we must have $\mathbb{S}(K, f) \subseteq R_{\mathbb{U}}(K)$ so since $R_{\mathbb{A}}(K)$ is the smallest of the $\mathbb{S}(K, g)$ the result follows. ■

In a way Theorem 19 is rather surprising in that one might initially have imagined that $R_{\mathbb{U}}(K)$, by its very definition, was about as specific a set as one could hope to describe. That $R_{\mathbb{A}}(K)$ can be strictly smaller than $R_{\mathbb{U}}(K)$ can be seen from the case when the $\sim_0$ equivalence classes look like

$$\{a_1, a_2\}, \{b_1, b_2\}, \{c_1, c_2\}, \{d_1, d_2, d_3, d_4\}.$$

In this case $R_{\mathbb{A}}$ gives $\{d_1, d_2, d_3, d_4\}$ whereas $R_{\mathbb{U}}$ just gives the union of all these sets.

We now briefly consider the relationship between the Regulative Reasons and $R_{\mathbb{U}}$. Since the set

$$R_i(K) = \{\, f \in K \mid \forall\, g \in K, |f^{-1}(i)| \geq |g^{-1}(i)| \,\}.$$

is definable in $\mathcal{M}$ $R_i(K)$ *is* a candidate for $R_{\mathbb{U}}(K)$. So if $|R_i(K)| < |R_{\mathbb{U}}(K)|$ it must be the case that there is another definable subset of $\mathcal{M}$ with the same size as $R_i(K)$. If $|R_i(K)| = |R_{\mathbb{U}}(K)|$ then in fact $R_i(K) = R_{\mathbb{U}}(K)$. From this point of view then $R_{\mathbb{U}}$ (and by Theorem 19 also $R_{\mathbb{A}}$) might be seen to be always at least as satisfactory as the R_i. On the other hand the R_i are in a practical sense computationally undemanding. [The computational complexity of the relation $f \sim_0 g$ between elements of K is currently unresolved, which strongly suggests that even if a polynominal time algorithm does exist it is far from transparent.]

We finally remark that, using the same examples as for $R_{\mathbb{A}}$, $R_{\mathbb{U}}$ also fails Obstinacy and Irrelevance.

6 An analogy with Game Theory

The situation we've been focussing on in this paper can be put quite naturally into game theoretic form. Although a full discussion of this, and the related reinterpretation of the results of the previous section in game theoretic terms are beyond the scope of this initial investigation on rationality-as-conformity, we nonetheless sketch here the main lines of this analysis.

The *conformity game*, as we might call it, is a two-person, non-cooperative game of complete yet imperfect information whose normal form goes like

this. Each agent is to choose one strategy out of a set of possible choices, identical for both agents. Each strategy corresponds to one element of $K = \{f_1, \ldots, f_k\}$, say. Agents get a (unique) positive payoff p if they play the same strategy, and nothing otherwise, all this being common knowledge. But since agents are to play simultaneously, they are clearly inaccessible to each other. Since it is a game of multiple Nash-equilibria, the conformity game is therefore a typical example of a (pure) *coordination game*, and as such, it is generally considered to be unsolvable within the framework of traditional game theory.

Recall that in section 2 we hinted at two general ways of solving the conformity game, corresponding to the following situations. Either worlds in K have no structure other than being distinct elements of a set, or worlds in K do have some structure, and in particular there are properties that might hold (be true) in (of) some worlds. In the former case we seem to be forced to accept that agents have no better way of playing the conformity game other than picking some world $f_i \in K$ at random (i.e. according to the uniform distribution). In the latter case, however, agents might use the information about the structure of the worlds in K to focus on some particularly distinguished possible world. On the assumption of like-mindedness, i.e. common reasoning, if one of those, say f_j should stand out as having some distinguished properties, agents will conclude that such properties are indeed intersubjectively accessible and hence select f_j. In the phraseology of coordination games those distinguished properties essentially contribute towards identifying a *salient* strategy, the corresponding equilibrium being called a *focal point*[8].

Though the analysis of the relation between rationality-as-conformity on the one hand and the selection of multiple Nash equilibria in (pure) coordination games on the other, goes beyond the scope of this paper, we note here that the Reasons we have been investigating in this paper qualify as natural candidates for a formalization of choice processes leading to focal points.

7 The rationality of conformity

The general results of this paper can be seen as formalizing the intuition according to which it would be irrational for two commonsensical agents to disagree *systematically* on their world view provided that it can be assumed that they are like-minded and that they are facing essentially the same choice

[8]The *loci classici* for pure coordination games and the related notions of salience and focal points are [26, 15]. Since then the literature on this topic developed enormously, yet particularly relevant to the present proposal are the more recent [17, 14, 11, 2, **?**].

problem. But why is this intuition reasonable within a general understanding of "rationality"?

There are surely several philosophical accounts of rationality that not only seem to be consistent with this intuition but seem to offer it some support. Nozick's theory of practical rationality [18] is surely to be included among those. According to the latter it is a sound principle of rationality agents that "sometimes accept something because others in our society do" ([18], p.129). This clearly finds it underpinnings in the intrinsic fallibility of human-level intelligent agents, indeed in the fact that within a society of rational agents, the systematic error of the majority is somehow less likely that the individual's. Moreover, as noted by Keynes, "Worldly wisdom teaches that it is better for reputation to fail conventionally than to succeed unconventionally" ([13] p.158).

The formalisation of rational choice behaviour we have pursued here is subtended by the assumption that reasons are devices agents apply to restrict their options, to go part, or sometimes even all, of the way to choosing a course of action or making a decision. Indeed, we have investigated choice behaviour as a two-stage process. In the first such stage agents apply reasons to *discard* those possible choices that are recognised as being unsuitable for the agent's purpose of conforming. If at the end of this process the agent is left with more than one *equally acceptable* option, then the actual choice is to be finalised by picking randomly (i.e. according to the uniform distribution) from this set.

Thus, it turns out that the general intuitions we have been following in the construction of the Regulative Reasons are remarkably close to those considered by Carnap (in the context of probabilistic confirmation theory) when developing his programme on Inductive Logic:

> The person X wishes to assign rational credence values to unknown propositions on the basis of the observations he has made. It is the purpose of inductive logic to help him to do that in a rational way; or, more precisely, to give him some general rules, each of which warns him against certain unreasonable steps. The rules do not in general lead him to specific values; they leave some freedom of choice within certain limits. What he chooses is not a credence value for a given proposition but rather certain features of a general policy for determining credence values. [8]

Indeed, as noted in passing above, the principles that make up the Regulative Reasons, are understood exactly as *policies* helping agents to achieve

their goal by forbidding them to undertake certain *unreasonable steps* that could prevent them from conforming.

Hence, again in consonance with the Carnapian perspective, we can go on and argue that our (idealised) modelling of Reasons does (ideally) also provide a justified definition of "rational" within the context, though what will be ultimately meant by this term is, like the scent of a rose, more easily felt than described.

Notice that neither Carnap's view nor the present account imply that whenever agents apply Reasons they will necessarily conform with probability 1. As we have seen the "rational choice" can simply be underdetermined with respect to the logical tools, the common reasoning, available to the agents, so that the possibility of disagreement in the final choice of a unique element from K cannot simply be ruled out in general (and indeed it would be rather exceptional for $R(K)$ to have size 1).

As a last point concerning the rationality of conformity, some illuminating suggestions can be found in the discussion of *radical interpretation* mainly championed by Davidson in a number of works (see the collections [3, 5]). The situation is one in which two agents are trying to establish successful communication despite their lack of a common language and without knowning anything about each other's view of the world.

According to Davidson, who inherits this intuition from Quine's analysis of radical translation ([24], ch.2), it is simply not possible for an agent to intepret successfully a speaker *without* assuming that she structures the worlds pretty much the same way the interpreter does. In other words, radical interpretation can only take place under an assumption which *mutatis mutandis* is entirely analogous to what we have been referring to here as like-mindedness: agents must assume that they share *common reasoning*. According to Davidson, this *Principle of Charity*, which ultimately provides fundamental clues about the others' cognitive make-up, is a necessary condition that agents must satisfy in order to solve a problem of radical interpretation, hence for establishing communication. But this amounts to activating the kind of structure that Davidson considers necessary in order for agents to be considered rational [4].

Again in connection with the Carnapian view, we note that according to the Principle of Charity agents should *discard* those possible interpretations that would make, to their eyes, the interpretee systematically wrong or (logically) inconsistent hence, yet without going into any of the subtleties of this topic, systematically irrational. In the formalization of Reasons this, as we have seen, amounts to discarding those possible worlds that are believed, on the fundamental assumption of their like-mindedness to prevent agents from converging or, in Davidson's felicitous terminology, *triangulating* on the same

possible world.

It is also interesting to notice here that a more or less implicit feature of (the radical) interpretation problems, which is shared by our rationality-as-conformity, consists in the fact that agents must share a common intention. Both the interpreter and the interpretee must in fact aim at assigning similar meanings to similar linguistic behaviours, that is, must aim at conforming[9].

As one might expect any ideas about the nature of rationality are likely to resonate with at least some of the multitude of viewpoints on the subject. The idea of rationality-as-conformity as we have presented it here is no exception and for this section we have just briefly noted some links with established positions on this matter. A fuller discussion may be found in the forthcoming [10].

8 Too many Reasons?

In this paper we have focussed on characterizing the choice process that would lead one isolated, common sensical, agent to conform to the behaviour of another like-minded yet inaccessible agent facing (essentially) the same choice problem. To this effect we have introduced what amount to four working Reasons, $R_0, R_1, R_{\mathbb{A}}, R_{\mathbb{U}}$. These arose through very different considerations. In the case of the Regulative Reasons through an adherence to rules, for $R_{\mathbb{A}}$ through an algorithm based on repeatedly trying to fulfill two desiderata, and for $R_{\mathbb{U}}$ through picking the smallest uniquely definable set within the given structure of the problem. This plurality of approaches and answers raises a vexing question.

How can we feel any confidence that there are not other approaches which will lead to entirely different answers?

As we have noted above, ideas and concepts from Game Theory would seem to have very definite application in generating Reasons. Furthermore similar hopes might be extended to other areas of mathematics, for example Social Choice Theory, which we have already alluded to in passing, Group Theory (the construction of $R_{\mathbb{A}}(K)$ could be seen as simply talking about permutation subgroups), Model Theory with its interests in definable subsets and Kolmogorov Complexity, with its emphasis on minimum description length. In short the answer to the question which headed this paragraph is that we can have little such confidence beyond the modicum which comes from having failed to find any ourselves.

[9]We touch upon some of these intriguing connection between radical interpretation, rationality-as-conformity and coordination games in [9].

Moreover, even with the candidates we already do have we have seen, from the examples given, that both the Regulative and Minimum Ambiguity Reasons appear capable, on their day, of monopolizing the right answer, the 'common sense' answer. Does that mean that even in this very simple context (let alone in the real world) there can be multiple common sense arguments? Or does it mean that we should try them all and pick the 'best answer'? (though that might seem to land us right back with the sort of problems we set out to answer in the first place!)

One advantage however that we should mention that the Minimum Ambiguity Reason and the Smallest Uniquely Definable Reason would appear to have over the Regulative Reasons is that they are easily generalizable. In place of permutations of K, equivalently automorphisms of $\mathcal{M}$, we take all automorphisms of the structure given to the agents and then define the corresponding $R_{\mathbb{A}}$ and $R_{\mathbb{U}}$ exactly analogously to the way we have here. To take a particular example if at the very start we had said that agents might not only receive the matrix with the rows and columns permuted but also possibly with 0 and 1 transposed then the natural structure would have been $\mathcal{M}$ with the constants 0 and 1 removed, i.e.

$$\langle \{0,1\} \cup A \cup K, \{0,1\}, A, K, =, Comp \rangle.$$

In this case an automorphism j corresponds to a permutation σ of A and a 1-1 function $\delta : \{0,1\} \longmapsto \{0,1\}$ such that

$$j \restriction \{0,1\} = \delta, \quad j \restriction A = \sigma^{-1}, \quad j(f) = \delta f \sigma \text{ for } f \in K.$$

Again the corresponding $R_{\mathbb{A}}$ and $R_{\mathbb{U}}$ can be defined and give in general practically worthwhile answers (i.e. non-trivial). However with this change the requirement of Renaming cannot be strengthened to what is expected here, i.e.

$$\delta R(K)\sigma = R(\delta K \sigma)$$

without reducing the possible Regulative Reasons to the trivial one alone – as can be seen by considering the initial step in the proof of Theorem 1.

9 Acknowledgements

We would like to thank the editor for his patience and encouragment and the referees whose unrestrained comments certainly resulted in this being a better paper.

References

[1] M. Aizermann and A. Malishevski. General theory of best variants choice: Some aspects. *IEEE Transactions on Automatic Control*, 26:1030–1040, 1981.

[2] C. Camerer. *Behavioral Game Theory: Experiments on Strategic Interaction*. Princeton, 2003.

[3] D. Davidson. *Inquiries into Truth and Interpretation*. Oxford University Press, 1984.

[4] D. Davidson. Rational Animals. In *Subjective, Intersubjective, Objective*, pages 95–105. Oxford University Press, 2001.

[5] D. Davidson. *Subjective, Intersubjective, Objective*. Oxford University Press, 2001.

[6] R. Elio, editor. *Commonsense, Reasoning, and Rationality*. New York: Oxford University Press, 2002.

[7] J. Halpern. *Reasoning About Uncertainty*. MIT Press, 2003.

[8] R. Hilpinen. Carnap's new system of inductive logic. *Synthese*, 25:307–333, 1973.

[9] H. Hosni. Conformity and interpretation. In *Prague International Colloquium on Logic, Games and Philosophy: Foundational Perspectives*, Prague, Czech Republic, October 2004.

[10] H. Hosni. *Doctoral Thesis*. School of Mathematics, The University of Manchester, Manchester, UK, 2005. http://www.maths.man.ac.uk/∼hykel/.

[11] M. Janssen. Rationalizing focal points. *Theory and Decision*, 50:119–148, 2001.

[12] G. Kalai, A. Rubinstein, and R. Spiegler. Rationalizing choice functions by multiple rationales. *Econometrica*, 70(6):2481–2488, 2002.

[13] J.M. Keynes. *The General Theory of Employment Interest and Money*. McMillan, London, (1936), 1951.

[14] S. Kraus, J. S. Rosenschein, and M. Fenster. Exploiting focal points among alternative solutions: Two approaches. *Annals of Mathematics and Artificial Intelligence*, 28(1-4):187–258, 2000.

[15] D. Lewis. *Convention: A philosophical study.* Harvard University Press, 1969.

[16] D. Marker. *Model theory: An Introduction.* Graduate Texts in Mathematics 217. Springer, 2002.

[17] J. Mehta, C. Strarmer, and Sugden R. The nature of salience: An experimental investigation of pure coordination. *The Americal Economic Review*, 84(3):658–673, 1994.

[18] R. Nozick. *The Nature of Rationality.* Princeton University Press, Princeton, 1993.

[19] J.B Paris. Common sense and maximum entropy. *Synthese*, 117(1):73–93, 1999.

[20] J.B. Paris and A. Vencovská. A note on the inevitability of maximum entropy. *International Journal of Approximated Reasoning*, 4:183–224, 1990.

[21] J.B. Paris and A. Vencovská. In defence of the maximum entropy inference process. *International Journal of Approximate Reasoning*, 17:77–103, 1997.

[22] J.B. Paris and A. Vencovská. Common sense and stochastic independence. In D. Corfield and J. Williamson, editors, *Foundations of Bayesianism*, pages 203–240. Kluwer Academic Press, 2001.

[23] C.R. Plott. Path independence, rationality and social choice. *Econometrica*, 41(6):1075–1091, 1973.

[24] W.V. Quine. *Word and Object.* MIT Press, Cambridge, Massachusetts, 1960.

[25] H. Rott. *Change, Choice and Inference : A Study of Belief Revision and Nonmonotonic Reasoning.* Oxford University Press, 2001.

[26] T. Schelling. *The strategy of conflict.* Harvard University Press, 1960.