Variable-Frame Level-N Theory*

Michael Bacharach
Institute of Economics and Statistics
University of Oxford

Dale Stahl
Deartment of Economics
University of Texas at Austin

27th November 1997

*This paper was written as part of the Research Programme 'Framing, Salience and Product Images' under ESRC Contract L122251002

Abstract

This paper develops a boundedly rational version of variable frame theory (VFT) of Bacharach [B93] by merging the variable frame conceptual apparatus with the level-n theory (LNT) of Stahl and Wilson [SW95]. A fundamental feature of VFT is that a player's options are determined by her 'frame': one 'option' is associated to each cell of a frame-induced partition of the actions. Coupled with the principles of payoff dominance (PD) and symmetry disqualification (SD), VFT resolves game-theoretic indeterminacy in pure coordination games, and the solution it defines explains the long puzzling 'Schelling competence', people's ability to solve coordination problems by collectively alighting on a salient coordination point.

The resolving power of VFT depends heavily on the refinement principles, PD and SD, which are not well supported in experiments. Further, in a theory which aspires to explain and predict real play, full rationality may be unduly strong. Both these points motivate this paper, in which we seek to apply the basic ideas of VFT to the boundedly rational LNT.

In our variable frame level-n theory, VFLNT, we retain the LNT notion that a level-0 player has no model of her coplayers, and that she randomizes uniformly over her options, but we define her options with respect to her frame. Then, an extended principle of insufficient reason uniquely defines level-0 behaviour. This model of their behaviour yields limited Schelling competence.

A level-1 player, unlike level-0s, is a rational Bayesian decision maker, and has a model of her coplayer. From this she determines a Bayes-optimal choice. In particular, she believes her coplayer to be level-0 (as in LNT). In our VFLNT, we further assume that a level-1 player has a frame f, and believes that her level-0 coplayer has a frame that is a subset of f. Level-2 players are very much like level-1 players except that they think their coplayer is either level-0 or level-1.

Like rational VFT, the VFLNT predicts the Schelling competence, but for quite different reasons. The VFLNT explanation combines a non-rational tendency at level-0 with rational capitalization on it by higher-level players. Thus, the resolution of coordination indeterminancies offered by rational VFT is more a product of the variable frame apparatus than of the rationality assumptions.

While for many games, rational VFT and VFLNT have qualitatively similar behavioural predictions, there are empirically testable differences. We propose a method of empirically testing this VFLNT.

1 Introduction

A Rational VFT

Variable frame theory (VFT) was introduced in [B93] as an alternative foundation for and refinement of Nash equilibrium. Bacharach and Bernasconi [BB97] applied VFT to experimental pure coordination games and found considerable support for the general approach, but at this point in time the evidence for the ancillary rationality assumptions of [BB97] remains mixed. The purpose of this paper is to develop a boundedly rational version of VFT by merging the variable-frame conceptual apparatus with the level-n theory (LNT) of [SW95].

We shall confine ourselves to object-choosing games, in which each player must choose one object from a finite set, and payoffs depend solely on which objects are chosen. This class of games is very wide: for example, we may consider different monetary amounts demanded in a bargaining game as different objects. Indeed any game with a common strategy set can be formally regarded as an object-choosing game.

A particularly simple class of object-choosing games are *matching games* ([B93], [S95], [J96]). A matching game is an object-choosing game in which, for each $a \in A$, each player receives the amount $y_a > 0$ if all choose a, and nothing otherwise.

A *frame* on A is a collection of families of characteristics of the objects of A that can be used to discriminate one from another, and to agglomerate some into groups of like kind. Given a set A, frames on A may vary in discriminating power. For example, if A is a set of physical objects which differ in terms of colour and shape, then there are several possible frames: both colour and shape might be used to discriminate the objects, or just colour, or just shape, or neither. If A consists of a red cube and three balls, one blue, one yellow, and one red, then the shape frame, the colour frame and the composite shape/colour frame partition A as {cube, balls}, as {red, not red}, and as {red cube, blue ball, yellow ball, red ball} respectively. Finally, we define the 'empty frame', which discriminates none of the objects from each other, and induces the trivial partition {objects}. We will call the elements of a partition 'cells'. A frame made up of a subset of the families of a given frame we will call a 'subframe' of the latter: for instance, the colour frame is a subframe of the shape/colour frame.

A fundamental feature of VFT is that a player's options are determined by her frame — specifically by the cells of the partitions of A induced by her frame and its subframes. The underlying principle is that to choose an object a player must have some way of 'mentally fixing' that object. This is usually a definite description, such as 'the yellow ball'.[1] If this were the only possibility, then if a player had the shape frame in our four-object example there would be only one object she could choose: the cube. However, VFT assumes that her ability to categorize the balls as balls gives her a second option, which consists in choosing at random, or 'picking', one of the three balls. Similarly, if she has the empty frame and only sees the objects of A as objects, she has the option of 'picking' any object. Thus, to each cell of a frame-induced partition is associated one 'option'.

[1]But it could be a 'nonconceptual representation' such as a certain ineffable 'look' or smell.

Example 1: Bottles (a 'simple Schelling game').[2] There are K bottles ($K \geq 3$), all identical except that one is hock-shaped, and the rest are claret-shaped (*Fig. 1*). Two players indicate their choice for an object simultaneously and privately. If both players choose the same bottle, then each receives a payoff of 1, and nothing otherwise.

[*Figure 1. Bottles*]

The shape frame partitions the bottles into two cells {hock-shaped, claret-shaped}, and hence induces two options: choosing the hock bottle (H), and picking a claret bottle (C). The empty frame (a subframe of the shape frame) induces a third option of picking any bottle (B). Thus for a player having the shape frame the option-induced payoff function is

	H	C	B
H	1	0	$1/K$
C	0	$1/(K-1)$	$1/K$
B	$1/K$	$1/K$	$1/K$

There are three pure-strategy Nash equilibria in options: HH, CC, and BB; in addition, there is a continuum of mixed strategy Nash equilibria, but these are not relevant to rational VFT. Payoff Dominance (PD), an auxiliary axiom of rational VFT [B93], selects the HH equilibrium, which is the Schelling solution.

Example 2: Coloured Bottles (a 'discrete Schelling game').. There are K bottles ($K \geq 3$), each a different colour, one hock-shaped, the rest claret-shaped, otherwise identical.

[*Figure 2. Coloured Bottles*]

In this example, colour (if noticed) generates an enriched frame, and this colour/shape frame induces the discrete partition of the objects, and hence K additional options of the form 'choose the ...-coloured bottle'. (Since the shape frame is a subframe of the colour/shape frame, the player still has the shaped-induced options.) We now have K additional pure-strategy Nash equilibria, all of which give an expected payoff of 1, like HH. To resolve this multiplicity, [B93], [J96] and [BB97] invoke the principle of Symmetry Disqualification (SD), which states that if two cells of a frame-induced partition contain the same number of objects then, since neither can be rationally presumed more probable than the other, any Nash equilibrium which contains the options given by these cells is disqualified as a solution to the game.

In Example 2, all the cells of the colour-induced partition contain exactly one object, and so all the K pure-strategy Nash equilibria in which each player chooses the same coloured bottle are symmetry-disqualified as solutions. Note that while the hock bottle is also in a singleton cell, it is induced by the shape frame, not the colour frame, and so is not disqualified.

What, the reader might ask, if one bottle is reddish and the others are different (but distinct) shades of green? Would SD still be applied? The answer is yes with respect to

[2] A 'simple Schelling game' [BB97] is a matching game in which there is just one family of characteristics for classifying the objects, and this family defines a single 'oddity' (an object which is the only one of its kind). Later we shall also meet 'discrete Schelling games', in which there is a second family of characteristics which discretely partitions the objects, and 'hi-lo Schelling games' in which the payoff for matching varies over objects.

the colour-induced discrete partition. However, now another frame is possible, namely {reddish, greenish}. This frame would lead to (reddish, reddish) as a Nash equilibrium, and this equilibrium would not be disqualified.

This discrete Schelling game example raises the issue of what frames are possible for individual players. In any collection of real bottles there are bound to be subtle differences among so-called identical bottles, not all of which are perceived or considered significant by every individual. Thus, different individuals may have different frames according to their discrimination abilities and their assessment of significance.

Recognizing the potential for individual differences, rational players must also have beliefs about the other players' discrimination faculties and hence their frames. Rational VFT imposes a restriction on these beliefs: the No Refinement principle (NR) that the belief-spaces considered possible for another player cannot refine a player's own belief-space. For example, if a player is not aware of any shape differences in Example 1, then she cannot have a belief about the other players' shape-induced options. For to have such a belief it is necessary that she be aware of shape differences.

Not only does VFT resolve game-theoretic indeterminacy in the above games, but the solution it defines also explains the long puzzling 'Schelling competence'. This is the ability to solve coordination problems such as Rendezvous by collectively alighting on a salient equilibrium or 'focal point' ([S60], [BB97], [M96]). In the traditional analysis of Example 1, there is one equilibrium for each bottle (in which both players choose that bottle). Intuitively, one of these, in which both choose the hock bottle, stands out, at least for players who notice the shapes; traditional theory fails to predict that shape-noticers will choose the hock bottle, but VFT does. In Example 2 also, going for the hock bottle seems, intuitively, salient, and once again traditional theory offers no prediction while VFT predicts that shape-noticers choose it.

B Derationalizing VFT

The resolving power of VFT in Examples 1 and 2 depends heavily on the refinement principles PD and SD. These are putative principles of rational behaviour whose claims are not secure. Although PD is very reasonable in pure coordination games, there is evidence from mixed-motive games (e.g. [HBB90]) that when pitted against Risk Dominance it often comes out on bottom; and the literature contains as yet no compelling *a priori* arguments for it. SD is also plausible, but it has as yet received no serious empirical support in 'hard cases'. Moreover, they are both principles of rational choice. So too is the Nash equilibrium principle itself which they refine. But in a theory which aspires to explain and predict real play, the notion of rationality which these principles express may be unduly strong.

Both these points — the unconfirmed status of PD and SD, and VFT's full-rationality approach to modelling the players — motivate what follows, in which we seek to apply the basic ideas of VFT to boundedly rational level-n theory (LNT) of [SW95]. The resulting theory we shall call the 'variable frame level-n theory' (VFLNT). It contains neither PD nor SD, and in it Nash equilibrium is a principle which describes only one specific type of player among several.

It might be wondered how much rationality is involved in the *variable frame* element of VFT, which we retain. The answer is, we think, only an irreducible minimum. Every player, however limited her reasoning powers, is a concept-possessor; otherwise, she could not understand her situation at all. And as long as players are concept-possessors, and there are alternative concepts which they might use in representing a particular choice situation, VFT necessarily applies.

True, there is *scope* for more rationality than this in framing a decision problem. In particular, it may be that agents engage in an optimized search activity in their framing, actively seeking richer or more convenient representations [NS72]. This is a research direction in VFT which we pursue elsewhere.

C The Games

In Section 2 we will review the rational VFT model, and in Section 3 we will present VFLNT and illustrate it by means of a few simple games. These are coordination games of the kinds studied in [S60], [B93], [S95], [J96] and [BB97]. It is important to consider such games, for it is an important test of the bounded rationality version of VFT that it should share the predictive success of the full rationality version for such games presented in [B93] and [BB97].

We shall also see, in Section 4, how VFLNT works when applied to games which are not pure coordination games, and see that in these games too it produces determinate solutions where traditional models do not.

2 A Review of VFT

Formally, we have a finite set A of objects, and a set $\mathcal{D}$ of families of characteristics of members of A. A *frame* is a subset of $\mathcal{D}$. Let $\mathcal{F}$ denote the set of frames. The empty frame represents the possibility that a player cannot classify the objects of A in any way (when her only option will be to pick in A).

Each family $d \in \mathcal{D}$ induces a partition P_d of A in the obvious way.. Similarly, with each frame $f \in \mathcal{F}$ we can associate a partition P_f: namely, the *join* of the P_d over $d \in f$. P_f is the cross-classification of the objects by all the families in f.

We take the distribution of frames in the population of players to be exogenous, and let $v(f)$ represent that distribution.

The frame-conditioned options are defined as follows. Let $F(f)$ denote the set of subsets of f. Then, for each $f' \in F(f)$, there is an option for each cell in $P_{f'}$, namely: if the cell contains a unique object, choosing that object, and otherwise picking an object from the cell (choosing one at random). Let Opt_f denote the complete set of options so defined: $Opt_f = \bigcup_{f' \in F(f)} P_{f'}$. Let $Opt = \bigcup_{f \in \mathcal{F}} Opt_f$.

A *decision rule* in VFT is a function, ϕ, which maps each possible frame f into an element of Opt_f, the set of options given by that frame. Starting with a payoff function for player i, u_i, which maps from $A \times A$ into the reals, it is straightforward to derive a payoff function U_i which maps from $Opt \times Opt$ into the reals.

In addition to a frame, a rational VFT player must have beliefs about the frame of her coplayer. Let $\widehat{v}(f' \mid f)$ represent a VFT player's belief that her coplayer has frame

f' conditional on herself having frame f. The NR principle prohibits a player from putting any positive weight on any frame that is not a subset of her own frame; that is,

(NR) the support of $\widehat{v}(* \mid f)$ is a member of $F(f)$.

The NR principle is a necessary implication of the very meaning of a frame. If a player believed it possible that another player was thinking about the shapes of the bottles, she too would *ipso facto* be aware of their shapes, and so S would be in her frame, contrary to hypothesis. Rational expectations require that $\widehat{v}(* \mid f)$ be formed from the true exogenous distribution v, by truncating v to $F(f)$ and rescaling.

Given a decision rule for the coplayer and subjective probabilities for the coplayer's frame conditional on the player's frame f, the expected utility of each option in Opt_f is well-defined. Specifically, for any option $o' \in Opt_f$,

$$EU(o' \mid f) = \sum_{f' \in F(f)} \sum_{o'' \in Opt_{f'}} U(o', o'')\, \widehat{v}(f' \mid f).$$

Then (ϕ, ϕ) is a *variable frame equilibrium* (vfe) if, for each f, $\phi(f)$ maximizes the player's expected utility against $\phi(*)$.

VFT further imposes the two following principles, Payoff Dominance and Symmetry Disqualification:

(PD) If a vfe is strictly payoff dominated by another vfe, then the payoff-dominated vfe is disqualified as a solution.[3]

(SD) If two cells of a frame-induced partition contain the same number of objects, any vfe which does not assign equal probability to these cells is disqualified as a solution.

A vfe that satisfies PD and SD is called a 'variable frame solution'. Despite these additional refinement principles, a game could have multiple variable frame solutions, though matching games generically have unique variable frame solutions [J96].

To illustrate, consider Example 1 in detail. $\mathcal{D} = \{\text{S}\}$, where S denotes the shape family. Hence, $\mathcal{F} = \{\emptyset, \text{S}\}$. If $f = \emptyset$, the only option is B, so trivially $\phi(\emptyset) = \text{B}$. If $f =$ S, then beliefs about frames become potentially important.. To determine whether ϕ is a vfe, we ask whether, if she hypothesizes that her coplayer uses ϕ, she maximizes her expected utility by conforming to ϕ. Using the vfe just determined for $\emptyset$, the relevant expected utilities are

$$\begin{aligned}
EU(H \mid S) &= v(S \mid S)\left[\phi(H \mid S) + \phi(B \mid S)/K\right] + \left[1 - v(S \mid S)\right]/K, \\
EU(C \mid S) &= v(S \mid S)\left[\phi(C \mid S)/(K-1) + \phi(B \mid S)/K\right] + \left[1 - v(S \mid S)\right]/K, \\
EU(B \mid S) &= 1/K,
\end{aligned}$$

where $\phi(* \mid S)$ is the indicator function defined by: $\phi(o \mid S) = 1$ if $\phi(S) = o$ and $\phi(o \mid S) = 0$ if $\phi(S) \neq o$. There are three vfe's, but PD selects HH as the unique

[3] Payoff-dominance is computed in the 'full' frame, $\mathcal{D}$.

variable frame solution given $v(S \mid S) > 0$ (i.e. provided our rational VFT player does not believe with certainty that her coplayer cannot perceive the shape difference).[4]

While SD played no role in Example 1, as we discussed in the Introduction, it is crucial to the VFT prediction for Example 2 (Coloured Bottles).

3 Variable Frame Level-n Theory

We build upon the frame apparatus of VFT. The basic use to which frames are put in VFLNT is in defining the options and beliefs a player works with in confronting her choice problems. Accordingly, a player's frame is the set of families of characteristics that she is *consciously aware of and pays attention to* as she contemplates a problem. It is often natural to identify these as the families of characteristics she *notices*. In a matching game in which the objects are physical and differ in the dimensions of colour C and shape S, a player might notice both, in which case she would have the CS (colour/shape) frame, or she might notice only colour, in which case she would have the C frame.

It is not sufficient that a player have the perceptual/conceptual ability to become aware of some family of characteristics; it is necessary that she pay conscious attention to those characteristics when contemplating her choice problem. There is ample psychological evidence that humans subconsciously filter out much of the information available to them as a means of coping with information overload (see e.g. [B82], [W1874]).[5] So it is probable that while an individual could distinguish each claret-shaped bottle if pressed to do so by specific questions or tasks, in a simple choice task these very subtle atomizing characteristics would be ignored. The outcome would be as if the individual could not perceive them.

A Level-0 Players

In LNT [SW95], a level-0 (L0) player has no model of her coplayers that induces a belief and an optimal choice against this belief, but instead a L0 player randomly picks an option in ignorance and without bias. In [SW95], a L0 player's behaviour is described by a uniform distribution over A. (From the viewpoint of the present model, [SW95] assumes implicitly that the L0 player's frame induces the discrete partition of A; this is indeed plausible, given the payoff-differences among the strategies.) A direct application of LNT to Example 1 would predict a uniform distribution of choices for L0 players, and this would also imply a uniform distribution of choices for all higher level players. In other words, LNT does not predict the Schelling competence.

In our VFLNT we retain the notion that a L0 player has no model of her coplayers, and that she randomizes uniformly over her options, but we define her options explicitly

[4] If $v(S \mid S) = 0$, then all three vfe's are variable frame solutions.

[5] Wundt wrote: "The basic phenomenon of all intellectual achievement is the so-called concentration of attention. It is understandable that in the appraisal of this phenomenon we attach importance first and therefore too exclusively to its positive side, to the grasping and clarification of certain presentations. But for the physiological appraisal it is clear that it is the negative side, the inhibition of the inflow of all other disturbing excitations ... which is most important."

with respect to her frame. Her frame f induces a partition P_f of A, and to each cell of this partition is associated an option: picking one of the objects in the cell or, if the cell has only one object, choosing that object. It is over these frame-induced options that the L0 player randomizes.[6]

In the rationalistic VFT described in [BB97], when a player has a frame f consisting of several families of characteristics, the player entertains the possibility that her coplayers might have a different frame, subject to NR, which restricts the possible frames of her coplayers to subsets of f. But this is a reasoning process about the coplayers and violates the notion of a L0 player. An L0 player has no ideas at all about her coplayers' frames.

The next step in developing VFLNT is to define a rule that for each potential frame f assigns a unique probability measure on A, which gives the probabilities that a L0 player will choose each object $a \in A$. Let $p_0 (a \mid f)$ denote the conditional probability of an L0 player's choosing object $a \in A$ when her frame is f. Our rule is derived from the following principles applied to the partition P_f. If P_f is the discrete partition, the principle of insufficient reason dictates that $p_0 (* \mid f)$ is the uniform distribution on A: the player must 'pick' an object from A. Extending the principle to a non-singleton cell of P_f, each object in the cell is equally probable (Object Picking); similarly, each cell in P_f is equally probable (Cell Picking). Thus the extended principle of insufficient reason uniquely defines $p_0 (* \mid f)$.

Since different individual players can have different perceptual/conceptual abilities and different subconscious information processes, they can have different frames. Given a set $\mathcal{D}$ of families of characteristics of the objects in A, let $\mathcal{F}$ denote the class of all subsets of $\mathcal{D}$, i.e. the class of all possible frames. Let $w(*)$ be a probability measure on $\mathcal{F}$, such that $w(f)$ represents the probability that a randomly selected L0 player has frame f. Then the population distribution of L0 players' choices over objects of A is given by

$$(1) \quad p_0 (*; w) = \sum_{f \in \mathcal{F}} p_0 (* \mid f)\, w(f) \,.$$

Even though L0 players are not Bayesian rational decision makers, this model of their behaviour yields limited Schelling competence.

Example 1, continued. Consider again the simple Schelling game Bottles. We can model $\mathcal{D}$ without loss of generality as consisting of two families: $\mathcal{D} = \{O, S\}$, where S is shape and O is some (possibly idiosyncratic) atomizing family of characteristics. For simplicity we assume that frame is never empty, so that $\mathcal{F} = \{O, S, OS\}$.[7] If the player notices the shape difference and ignores idiosyncratic differences, $f =$ S, and P_f is {hock-shaped, claret-shaped}. Therefore, $p_0 (\mathit{hockbottle} \mid S) = 1/2$, and $p_0 (a \mid S) = 1/2\,(K-1)$ whenever a is claret-shaped. On the other hand, if the player does not

[6] In Example 1, it is possible that an L0 player initially notices subtle imperfections in the claret-shaped bottles, but subsequently ignores them and defines her options as {choose the hock-shape, pick a claret-shape}. Rather than interpret this as the L0 player using a subframe, we prefer to interpret the first impressions of glass imperfections as subconscious perceptions that are screened out, leaving shape as the family she is aware of and uses in defining her options, hence her frame.

[7] Without loss of potential behaviours, we have ignored the $\emptyset$ frame, since it would merely lead to uniformly random choices for L0 and L1 players, and such behaviour is already possibly with O present.

notice the shape difference, $f = \mathrm{O}$, and P_f is the discrete partition, so $p_0(a \mid O) = 1/K$ for all $a \in A$. Finally, if the player pays attention to both shape and idiosyncratic difference, $f = \mathrm{OS}$, and P_f is the discrete partition, so $p_0(a \mid OS) = 1/K$ for all $a \in A$.[8]

For notational convenience, label the hock bottle 1 and the others $2, \ldots, K$. Then, the aggregate distribution of L0 choices is

$$
\begin{aligned}
(2) \quad p_0(1; w) &= [w(S)/2] + [1 - w(S)]/K = (1/K) + w(S)(K-2)/2K, \\
p_0(a; w) &= [w(S)/2(K-1)] + [1 - w(S)]/K \\
&= (1/K) - w(S)(K-2)/2K(K-1) \qquad (a \neq 1)
\end{aligned}
$$

Thus, if $w(S) > 0$, L0 players choose the hock bottle nearly half the time when K is large, and as long as $w(S) > 0$ it is the modal choice among L0 players. In this sense, the L0 players possess a limited Schelling competence. We shall see later that this tilt in the choice distribution of a L0 towards the hock bottle, though it may be slight, can induce a much greater tilt in the choices of higher-level players.

Example 2, continued: Coloured Bottles. The L0 behavioural predictions for this discrete Schelling game are exactly the same as in Example 1. To see this, note that the colour family induces the discrete partition, and hence the same behaviour as the idiosyncratic atomizing family O. Thus, the only probability parameter that matters behaviourally is $w(S)$, the probability that the L0 player uses the S family exclusively. Of course, it is not unreasonable to expect the actual value of $w(S)$ in Coloured Bottles to be less than in the previous example, because using the S frame exclusively entails ignoring both the idiosyncratic differences and the colour differences. Nonetheless, as long as $w(S) > 0$, the hock bottle is more likely to be chosen by L0 players than any claret bottle, a prediction that VFT makes only after imposing Symmetry Disqualification.

B Level-1 Players

A level-1 player, unlike L0s, is a rational Bayesian decision-maker, and as such she has a model of her coplayer that induces a belief from which she determines an optimal choice. In particular, she believes him to be L0 (as in LNT) and that he behaves according to $p_0(* \mid f)$. In our VFLNT, we further assume that a L1 player has a frame f, and believes that her L0 coplayer has a frame that is a subset of f, as required by NR. Since a L0's frame may in fact contain dimensions not in the L1's frame, the NR principle implies that the L1 may be mistaken. She may also be mistaken in believing that her coplayer is an L0; but this feature of L1s is inescapable, for much the same reason as the NR principle: if a L1 player were to conceive of a higher-level type and know what that meant, then *ipso facto* she would be able to do the higher-level reasoning.

We give precision to our assumptions about a L1's model of her coplayer as follows. We begin by supposing that at some (sub)conscious level, perhaps very briefly, the L1 player reacts psychologically like a L0 player, but then conscious reasoning takes over. During this 'L0 moment', a frame f is generated with probability $w(f)$, equal to that

[8] Notice that whenever an atomizing family O belongs to a L0 player's frame f, the partition P_f will be discrete, and so $p_0(* \mid f)$ will be uniform for all such f, and in this sense all such frames are equivalent, and hence could be represented by the $f = \mathrm{O}$ case.

for L0 players. Her own frame thus determined is a major input into the L1 belief. Now for an L1 who has frame f we need to specify the subjective probabilities for her coplayer's frame; these we will denote by $\widehat{w}\,(* \mid f)$. We make two assumptions.

The first simply applies the NR principle of VFT to L1 players: an L1 player cannot believe that her L0 coplayer has a frame that is not a subset of f. Let $F\,(f)$ denote the subsets of the families that define f; then we assume that the support of $\widehat{w}\,(* \mid f)$ is a member of $F\,(f)$.

Second, an L1's belief about *which* frame in $F(f)$ her coplayer has may be more or less biased towards her own frame, f. If the L1 player has rational expectations (subject to NR), then $\widehat{w}\,(* \mid f)$ will be equal to $w\,(*)$ truncated to $F\,(f)$. On the other hand, it is reasonable to expect that a L1's model of a L0 player will suffer from 'noticer bias': the tendency by an agent who has thought of a certain way of classifying objects to overestimate or underestimate the probability that others will classify them in the same way. ([BB97] found evidence of positive noticer bias.) It is, for example, quite possible that the L1 player will simply project her own state of mind onto the other player; i.e. she will believe that the other player has frame f also; i.e. $\widehat{w}\,(f \mid f) = 1$. But this is an extreme case. A simple form of positive noticer bias of variable degree can be captured by one parameter in the following manner. Let

$$
\begin{aligned}
(3) \qquad \widehat{w}\,(f' \mid f) &= \frac{b\,(f' \mid f)\, w\,(f')}{\sum_{f'' \in F(f)} b\,(f'' \mid f)\, w\,(f'')} \qquad for f' \in F\,(f)\,, \\
&= 0 \qquad otherwise,
\end{aligned}
$$

where

$$
b\,(f' \mid f) = b > 1 iff' = f and b\,(f' \mid f) = 1 otherwise.
$$

Thus formula (3) truncates $w\,(*)$ to the set of frames compatible with f and distorts it in the direction of positive noticer bias.

Continuing with the development of the VFLNT, we assume that a L1 player having frame f has probabilities for her coplayer's choice given by

$$
(4) \qquad q_1\,(a; \widehat{w} \mid f) = \sum_{f' \in F(f)} p_0\,(a \mid f')\, \widehat{w}\,(f' \mid f) \qquad (a \in A)\,.
$$

When our L1 player's frame f contains some atomizing family of characteristics, say O, then since P_f is the discrete partition equation, (4) is compatible with f. However, when P_f is not the discrete partition, then the interpretation of (4) compatible with f is not as obvious. We first interpret $p_0\,(a \mid f')$ from L1's viewpoint, and second $q_1\,(a; \widehat{w} \mid f)$. Let $P_f\,(a)$ denote the cell of P_f that contains object a. Then, by the NR principle, $P_{f'}$ is a coarsening of P_f, so statements about $P_{f'}\,(a)$ are clearly permissible. In particular, while the L1 player may not have individuating ways of thinking about each $a' \in P_{f'}\,(a)$ (not even a personal labelling), she can affirm the *quantified* statement: "Conditionally on my coplayer having frame f', for each $a' \in P_{f'}\,(a)$, the probability that he will choose it is $p_0\,(a' \mid f')$." Then, since $P_f\,(a) \subseteq P_{f'}\,(a)$, it follows that she can affirm the *quantified* statement: "For each $a \in P_f(a)$, the probability that he will choose it is $q_1(a'; \widehat{w} \mid f)$."[9]

We assume that a L1 player chooses an option that maximizes her perceived expected utility. Let $u_{aa'}$ denote the player's payoff when she chooses a and her coplayer chooses a'. Consider an arbitrary option $o \in Opt_f = \bigcup_{f' \in F(f)} P_{f'}$. If the cell which defines o contains a single object a, the player's perceived expected utility of o is

$$U(o; \widehat{w} \mid f) = \sum_{a' \in A} q_1(a'; \widehat{w} \mid f)\, u_{aa'}.$$

The fact that all objects that belong to a single option are equally likely to be chosen implies the general formula for the perceived expected utility of an option o

$$(5) \quad U(o; \widehat{w} \mid f) = \sum_{a \in o} \frac{1}{N(o)} \sum_{a' \in A} q_1(a'; \widehat{w} \mid f)\, u_{aa'},$$

where $N(o)$ is the size of o.

In the spirit of LNT we allow the L1 player to suffer from computational error, and so we posit a logit probabilistic choice function over options

$$(6) \quad \ell_1(o; \gamma_1, \widehat{w} \mid f) = \frac{\exp[\gamma_1 U(o; \widehat{w} \mid f)]}{\sum_{o' \in Opt_f} \exp[\gamma_1 U(o'; \widehat{w} \mid f)]} \qquad (o \in Opt_f),$$

where γ_1 is the precision parameter (see [SW95]).

The L1 player's choice probability for an object $a \in A$ is the sum of the conditional probabilities that she chooses a given that she chooses option o, over all o which contain a:

$$(7) \quad p_1(a; \widehat{w} \mid f) = \sum_{o \in Opt_f(a)} \frac{1}{N(o)} \ell_1(o; \gamma_1, \widehat{w} \mid f) \qquad (a \in A),$$

where $Opt_f(a)$ is the subset of Opt_f which contains a.

The population choice probability for L1 players is

$$(8) \quad p_1(a; \gamma_1, \widehat{w}, w) = \sum_{f \in \mathcal{F}} p_1(a; \widehat{w}, \gamma_1 \mid f)\, w(f) \qquad (a \in A).$$

Example 1, continued. In the simple Schelling games (Bottles), an L1 player can have three possible frames: O, S or OS. If a L1 player has frame O, then she believes her (L0) coplayer has frame O also, and so $\widehat{w}(S) = 0$, and from (4) above, $q_1(a; \widehat{w} \mid O) = 1/K$ for all $a \in A$. All choices are equally good, so $p_1(*; \gamma_1, \widehat{w} \mid O)$ is also uniform. On the other hand, if an L1 player has frame S, then she believes her (L0) coplayer has frame S also, and so $\widehat{w}(S) = 1$, and from (4) above, $q_1(1; \widehat{w} \mid S) = 1/2$ and $q_1(a; \widehat{w} \mid S) = 1/2(K-1)$ for $a \neq 1$. The best response is obviously to choose the hock bottle. For any positive precision ($\gamma_1 > 0$), the modal choice will be the hock bottle, and as the precision increases, $p_1(1; \gamma_1, \widehat{w}, w \mid S)$ approaches 1.

[9] This quantified reading recalls Sugden's notion of an 'existential game' [S95]. There the quantifier is existential rather than universal. In both cases the modelling of the player's conception of the situation involves quantifiers because she lacks direct perceptual or other knowledge of individual things — in our case objects, in his case strategies.

The more complex case is when a L1 player has frame SO, because then she will contemplate whether the L0 player has frame S or O or SO. The L1 player's probabilities for her coplayer's choices depend on $\widehat{w}(S \mid f)$, with $f = \text{SO}$, and are given by

$$\text{(9)} \quad \begin{aligned} q_1(1; \widehat{w} \mid f) &= (1/K) + \widehat{w}(S \mid f)(K-2)/2K, \\ q_1(a; \widehat{w} \mid f) &= (1/K) - \widehat{w}(S \mid f)(K-2)/2K(K-1) \qquad (a \neq 1). \end{aligned}$$

Choosing the hock bottle is still optimal, and hence will be the most likely choice of L1 players. The hock bottle belongs to two options: $Opt_{OS}(1) = \{1, hock - shaped\}$, where the second '1' denotes the idiosyncratic option whose only member is the hock bottle. Because these options contain just the same objects (each contains only the hock bottle), their perceived expected utilities $U(1; \widehat{w} \mid OS)$, $U(hock - shaped; \widehat{w} \mid OS)$ are equal, and so too are their option choice probabilities $\ell_1(1; \gamma_1, \widehat{w} \mid OS)$, $\ell_1(hock - shaped; \gamma_1, \widehat{w} \mid OS)$. The latter sum to the object choice probability $p_1(1; \widehat{w} \mid OS)$.

Note that the above equation for $q_1(*; \widehat{w} \mid f)$ holds for all f, since $\widehat{w}(S \mid S) = 1$ and $\widehat{w}(S \mid O) = 0$ by virtue of NR.

Example 2, continued. In Coloured Bottles, (9) still defines the prior of a L1 player, but now f ranges over $\mathcal{F} = \{C, O, S, CO, CS, OS, COS\}$. Note, however, that this extended set of frames generates the same set of partitions (and hence options) as $\mathcal{F}' = \{O', S, O'S\}$, where O$'$ stands for C and/or O, so one could restrict the range of f to $\mathcal{F}'$, thereby revealing that the only difference between L1 behaviour in Bottles and Coloured Bottles would stem from the difference, if any, between $\widehat{w}(S \mid OS)$ and $\widehat{w}(S \mid O'S)$.

C Level-2 Players

Level-2 (L2) types are very much like L1 types except that they think their coplayer is either L0 or L1. Our L2 model is motivated by the following story of iterative thinking. A L2 player begins with an 'L0 moment' during which a frame f is generated. Next, the L2 player starts thinking like a L1 player, forming beliefs $\widehat{w}(* \mid f)$ about L0 frames, which leads to a tentative probability measure $q_1(*; \widehat{w} \mid f)$. As our L2 player computes the expected utility to this tentative measure, she realizes that some of the other players could do the same. She forms a belief about the proportion of L0 types in the population, α_0, and she believes that everyone else is a L1 type. As part of this next step, the L2 player realizes that some of these L1 players might not have the same frame as her, but perhaps some subframe in $F(f)$. For each $f' \in F(f)$, the L2 player predicts that the L1 behaviour will be $p_1(*; \widehat{\gamma}_1, \widehat{w} \mid f')$, where $\widehat{\gamma}_1$ is the L2 player's point estimate of γ_1. Hence, the composite prior of the L2 player is

$$\text{(10)} \quad q_2(*; \widehat{\alpha}, \widehat{\gamma}, \widehat{w} \mid f) = \sum_{f' \in F(f)} \left[\widehat{\alpha}_0 p_0(* \mid f') + (1 - \widehat{\alpha}_0) p_1(*; \widehat{\gamma}_1, \widehat{w} \mid f') \widehat{w}(f' \mid f)\right],$$

where $\widehat{\alpha}_0$ is her point estimate of α_0, and $\widehat{w}(f' \mid f)$ is given by (3), just as for a L1 player. We assume that, since framing is prior to reasoning, not only the likelihood of frames but also beliefs about others' frames are common to all reasoning levels.

Thus, L2s have the same frame-conditioned model of L0s as L1s do. L2s, like L1s, may suffer from noticer bias in assessing their coplayer's likely frame; and this bias carries over into their model of how L1s think about L0's frame (that is, they ascribe to L1's their own point estimates of frame probabilities). Like L1s, they get everything right except for, possibly, two things: this bias, and the type distribution (i.e. $\widehat{\alpha}_0$ could be wrong, and there are in fact types other than L0s and L1s in the population).

Just as for a L1 player, we assume that a L2 player chooses an option (based on frame f) that maximizes her perceived expected utility. The perceived expected utility of an option o is now

$$(11)\quad U(o;\widehat{w} \mid f) = \sum_{a\in o} \frac{1}{N(o)} \sum_{a'\in A} q_2(a';\widehat{\alpha}_0,\widehat{\gamma}_1,\widehat{w} \mid f)\, u_{aa'}$$

and, as before allowing the L2 player to suffer from computational error, she has the logit probabilistic choice function over options

$$(12)\quad \ell_2(o;\gamma_2,\widehat{w} \mid f) = \frac{\exp[\gamma_2 U(o;\widehat{w} \mid f)]}{\sum_{o'\in Opt_f} \exp[\gamma_2 U(o';\widehat{w} \mid f)]} \qquad (o \in Opt_f),$$

where γ_2 is the precision parameter. As before, this leads to the object choice function

$$(13)\quad p_2(a;\widehat{w} \mid f) = \sum_{o\in Opt_f(a)} \frac{1}{N(o)} \ell_2(o;\gamma_2,\widehat{w} \mid f) \qquad (a \in A).$$

Thus, the L2 player is characterized by three parameters $\widehat{\alpha}_0, \widehat{\gamma}_1, \gamma_2$ and the common beliefs $\widehat{w}$. However, for symmetric games, because a low precision L1 player behaves like a L0 player, the L2 subjective probabilities can be reasonably approximated by

$$q_2(*;\widehat{\gamma}_1,\widehat{w} \mid f) = \sum_{f'\in F(f)} [p_1(*;\widehat{\gamma}_1,\widehat{w} \mid f')\,\widehat{w}(f' \mid f)],$$

which eliminates the parameter $\widehat{\alpha}_0$, allowing the parameter $\widehat{\gamma}_1$ to capture both the precision of L1 types and the presence of L0 types. [SW95] found that it would be difficult to identify both $\widehat{\alpha}_0$ and $\widehat{\gamma}_1$ from experimental data.

Example 1, continued: Bottles. Note that the presence of $p_1(* \mid f')$ in (10) tilts the L2 prior more in the direction of the hock bottle. This tilt will magnify the expected-utility difference between the hock bottle and the claret bottles, and the L2 player will therefore be even more likely to choose the hock bottle.

Example 2, continued: Coloured Bottles. For the same reasons, the L2 player will, in this example as well, be even more likely to choose the hock bottle.

D Higher Level-n Players

The general scheme for specifying a level-n ($n > 2$) type should be clear. For example, a level-3 type will involve a frame f, beliefs about others' frames $\widehat{w}(*)$, beliefs about the proportion of L0, L1 and L2 types $(\widehat{\alpha}_0, \widehat{\alpha}_1, 1-\widehat{\alpha}_0-\widehat{\alpha}_1)$, beliefs about the precision of lower types $(\widehat{\gamma}_1, \widehat{\gamma}_2)$, and a new precision parameter (γ_3). Thus, there are no new conceptual elements for higher level types. The assumption that beliefs about frames and noticer bias are independent of other cognitive reasoning processes allows us to formulate the model in terms of a common family of conditional probability measures over frames, $w(* \mid *)$, and a common noticer bias (b) implicit in $\widehat{w}(* \mid *)$.

However, we are sceptical about the need to continue this hierarchy levels above L2 in any empirical test of VFLNT, for the reasons given in [SW95]. On the other hand, in order to have an encompassing model for empirical testing, [SW95] extended the hierarchical LNT model to include the traditional Nash equilibrium type, and a 'worldly' type who thinks her coplayer is either a traditional Nash type or an L0 or L1 type. We take up a similar extension of the VFLNT in subsection G below, but first we present two interesting theoretical observations concerning VFLNT in its current form.

E VFLNT and the Schelling Competence

Like rational VFT, VFLNT predicts the Schelling competence, people's ability to solve coordination problems by collectively alighting on a *salient* coordination point. The root of the VFLNT explanation is that when there are obvious differences of some kind among the objects in A, there is a tendency for level-0 players to define their options by those differences. This tendency is unthinking. It acts as a 'seed' which propagates upwards through the hierarchy. The VFLNT explanation of the population tendency towards the focal point combines this non-rational tendency at level-0 with rational capitalization on it by higher-level players.

Thus VFLNT explains focal play in a population by asymmetric rationality in classification, which it adds to the asymmetric strategic rationality of LNT. VFLNT does not address the controversial question 'Does pure rationality indicate choosing the salient option when playing with a purely rational coplayer?' ([G75], [H78], [G89]); instead, it takes as its starting point the empirical existence of a hierarchy of levels of rationality, and asks how such a hierarchy behaves in games with salient options. It builds on the simple idea that if some people choose the salient object without thinking then it is rational to do so oneself. But it gives a precise formulation to this imprecise idea by incorporating the model of agents' descriptions provided by VFT.

F Contrasting VFT and VFLNT

While the examples so far suggest that the rational VFT and VFLNT have qualitatively similar behavioural predictions, there are empirically testable differences. To illustrate this, we introduce a 'hi-lo Schelling game', Bottles with Payoff Differences. First, we make precise the assumption that payoff differences are always noticed.

For any matching game, let Y be the family of characteristics of objects of A which puts a and a' in different cells if and only if matching on a and matching on a' give different payoffs. We assume that Y is an element of every player's frame.[10]

Example 3: Bottles with Payoff Differences. As in Example 1, there are K bottles ($K \geq 3$), but now a match on the hock bottle pays only $y \in (1/(K-1), 1)$, while a match on a claret bottle pays 1.

[10] This precludes the possibility that a L0 player's frame is the singleton O, the family of idiosyncratic atomizing characteristics. However, observe that as long as the L0 player's frame could be OY, since P_{OY} is the discrete partition the set of possible options is not reduced by our assumption, and the exposition is less cumbersome.

As before, we suppose there is an atomizing family of characteristics O in addition to the shape family S and Y; thus, $\mathcal{F} = \{Y, OY, SY, OSY\}$; note that we assume that every frame contains Y. Although we have introduced payoff differences (Y), this family of characteristics introduces no new options, since here Y partitions the objects exactly as shape (S) does. Therefore, without loss of generality, we can take the set of possible frames to be $\mathcal{F}' = \{S', OS'\}$, where S′ stands for S and/or Y.

Choosing the hock bottle is the unique VFT solution. To see this, first observe that a rational player must be able to compute expected utilities, and therefore Y is always an element of a player's frame in VFT. Consequently, $v\left(O \mid OS'\right) = \mathrm{O}$, since otherwise the coplayer would not be a rational player (could not compute expected utilities). Thus, when f = OS′, the support of coplayer subframes is $\mathcal{F}'$. It is now straightforward to verify that, given $y > 1/\left(K-1\right)$, payoff dominance (PD) always picks the {hock, hock} vfe.

In VFLNT, the aggregate L0 behaviour will be given by (2), with S′ replacing S. An L1 player's subjective probabilities for her coplayer's choice are given by (9), so her expected utility is

$$EU\left(1 \mid f\right) = \left[2 + \widehat{w}\left(K-2\right)\right] y/2K,$$

and

$$EU\left(a \mid f\right) = \left[2 - \widehat{w}\left(K-2\right)\right]/\left(K-1\right)/2K \qquad \left(a \neq 1\right),$$

where $\widehat{w} = \widehat{w}\left(S' \mid f\right)$ represents the L1 player's belief about the L0 frame, and f ranges over $\mathcal{F}'$.

When f = S′, then $\widehat{w} = 1$, so $EU_f\left(1\right) > EU_f\left(a\right)$ for all $a \neq 1$; i.e. choosing the hock bottle is best.

However, $\widehat{w} < 1$ is quite possible and reasonable when f = OS′. Then, choosing the hock bottle is best iff $y > W/(K-1)$, where

$$W = \left[2K - 2 - \left(K-2\right)\widehat{w}\right]/\left[\left(K-2\right)\widehat{w} + 2\right],$$

which exceeds 1 for all $\widehat{w} < 1$. Thus, for any $\widehat{w}\left(S' \mid OS'\right) < 1$, by setting y sufficiently close to $1/\left(K-1\right)$, picking a claret bottle will be best for an L1 player with frame OS′. E.g. if $K = 3$, and $y = 0.51$, the claret option is best for all $\widehat{w} < 0.97$. In this case, the L0 tilt towards the hock bottle is not sufficient to overcome the lower payoff. For all higher level-n players, the tilt of the L1 player away from the hock bottle will be magnified. Thus, for the composite OS′ frame, VFLNT predicts predominantly claret bottle choices, in contrast to VFT.

G E Players and Worldly Players

For the purpose of empirically testing VFLNT it is very useful to have a model that encompasses both the pure theory under consideration and alternative theories. Since VFLNT is essentially a mixture of probabilistic choice functions over types of players, it can easily be extended to include an alternative theory by introducing a type whose behaviour corresponds to the alternative theory.

To introduce a type corresponding to VFT, we define a 'type E' player to be one who conforms to the equilibrium theory VFT. When there is a unique variable frame solution ϕ a type E player will conform to it, choosing option $\phi(f)$ when in frame f. Hence, the aggregate distribution of object choices by type E players is given by

$$(14)\quad p_E(a; v \mid f) = \sum_{f\in\mathcal{F}} \left[\sum_{o\in Opt_f(a)} \frac{1}{N(o)} \phi(f) \right] v(f) \qquad (a \in A).$$

When there are multiple variable frame solutions, an E player conforms to some variable frame solution, and we assume that the population of type E players uniformly distribute themselves over all variable frame solutions.[11]

To test various aspects of VFT, such as symmetry disqualification, we would introduce another version of type E that conforms to VFT except for SD. To test the rational expectations assumption of VFT, we would introduce a version of type E in which $\widehat{v}(* \mid f)$ contains potential noticer bias, analogously to $\widehat{w}(* \mid f)$ in VFLNT.

Similarly to LNT, we define a 'worldly' type as a player who thinks her coplayer is either a type E player or else an L0 or L1 type. Let $\widehat{\alpha}_0$ and $\widehat{\alpha}_1$ denote a worldly type's subjective belief as to the proportion of the population of coplayers that are L0 and L1 types respectively; a proportion $1-\widehat{\alpha}_0-\widehat{\alpha}_1$ is believed to be type E. Then, a worldly player's probabilities for her coplayer's choices are given by

$$\begin{aligned}(15)\quad q_W(*; v, \widehat{w} \mid f) \;=\; & \sum_{f'\in F(f)} \left[\widehat{\alpha}_0 p_0(*; \widehat{w} \mid f') + \widehat{\alpha}_1 p_1(*; \widehat{w} \mid f')\right.\\ & \left. + (1-\widehat{\alpha}_0-\widehat{\alpha}_1)\, p_E(*; v \mid f)\right] v(f' \mid f).\end{aligned}$$

Note that since the meaning of 'having frame f' is different for type E players and Ln players, there is no *a priori* reason that a worldly type should believe that type E players and Ln players have the same distribution of frames; thus, (14) uses $\widehat{w}$ to denote the worldly type's belief about the distribution of Ln frames. While this $\widehat{w}$ has the same functional role as in pure VFLNT, it is not necessarily the same distribution: i.e., worldly types and Ln types could have different beliefs about the distribution of Ln frames. That they be the same is an hypothesis that could be empirically tested.

4 VFLNT and Non-Coordination Games

The games we have been discussing are all examples of object-choosing games, in which the effect of different frames is to classify a set of objects in different ways, so generating different option sets. This class of games is very wide, since 'objects' can be of all kinds, not necessarily physical: for example, we may consider different monetary amounts demanded, in a bargaining game, as different objects, frameable as fair-unfair, or by arithmetical properties, and so on. VFLNT can, moreover, easily be extended beyond object-choosing games, to cases where an action rather than an object is describable in different ways: for example, the choices in a Prisoner's Dilemma might be thought of as

[11] In lieu of this latter assumption, one could restrict attention to cases in which there is a unique variable frame solution.

cooperative-uncooperative, or trusting-doublecrossing, or as as giving your coplayer3-giving yourself 1 [EC66]. We do not, however, consider this extension in the present paper.

Notice that nothing in VFLNT requires the payoff structure to be coordinative rather than competitive or mixed. This is illustrated in the next example, which is of a competitive object-choosing game.

In a game studied by Rubinstein, Tversky and Heller [RTH95], a Hider puts an object in one of four boxes; one box is highly salient (the only one labelled Z) and one is of minimal salience (the only box labelled X to lack a salient position). The Hider receives 1 or 0 according as the Seeker does or does not look in the right box, and the game is zero-sum. The unique Nash equilibrium in a traditional analysis without framing effects is for both Hider and Seeker to randomize uniformly over A. Rubinstein *et al.* predicted that both Hiders and Seekers would favour the 'naive strategies' of hiding (and seeking) at random in the set of boxes of minimal salience; this choice was favoured to a significant degree by Hiders and more strongly by Seekers. These strategies are naive because they reveal 'very limited strategic reasoning on the part of the players'. If a Hider thinks that the nonsalient box is a good idea because the Seeker will be drawn to a salient one, she is treating the Seeker as a non-reasoner; she fails to credit the Seeker with the ability to anticipate her strategy. In effect, the Hider treats the Seeker as L0. What does VFLNT predict? The spirit of the Rubinstein *et al.* game may be captured by the following slightly simplified version.

Example 4: Hide and Seek. A consists of three boxes, one (1) marked with a Z and two (2, 3) with an X; one of the Xs, 3, is off-white and 1 and 2 are white. $\mathcal{D} = \{L, C\}$, where L denotes the family of letter labels, and C denotes the family of white colours. Hence, $\mathcal{F} = \{L, C, LC\}$.[12]

An L0 Seeker would have frame conditional choice probabilities

$$p_{0S}\left(* \mid L\right) = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right), \quad p_{0S}\left(* \mid C\right) = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}\right), \quad p_{0S}\left(* \mid LC\right) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right).$$

Then, aggregate behaviour of L0 Seekers would be

$$p_{0S}\left(* \mid L\right) = w_L\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right) + w_C\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}\right) + \left(1 - w_L - w_C\right)\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right).$$

It is reasonable to assume that $w_L > w_C$, in which case the likelihood of the choices are ordered as $p_{0S}\left(1\right) > p_{0S}\left(3\right) > p_{0S}\left(2\right)$. In other words, an L0 Seeker is more likely to look in the Z box and least likely to look in the plain white X box.

As first blush we might suppose that L0 Hiders behave similarly, hiding the object in the most salient box. However, such behaviour is exceedingly insensitive to the nature of the situation. Without having any hint of a model of Seekers, a non-thinking Hider is likely to have her attention drawn precisely to inconspicuous boxes (ones which lack 'first-order' salience). Evolutionary considerations suggest this, for early humans had frequent need to hide from predators, creating a natural selection for instincts to hide in inconspicuous places. In contrast, looking for food, water and shelter in the

[12] Since P_{LC} is the discrete partition, picking any box is already an option for the L0 players, so ignoring the $\emptyset$ frame has no practical consequence, and simplifies the exposition.

most conspicuous places would have been conspicuously rewarded.[13] All this can be accommodated within the variable frame approach. Variable frame theory makes no assumption about the relationship between the availability functions of players with asymmetric roles. There is no reason why players who have the tasks of 'looking in a box' and 'shutting up in a box' should be struck by the same characteristics of boxes.

Accordingly, we hypothesize that the instinctive L0 response is anti-salient. Specifically, we suppose that the probabilistic choice of an L0 Hider with frame f is

$$p_{0H}(f) = 2\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) - p_{0S}(f).$$

In other words, we reflect the L0 Seeker's behaviour through the uniform distribution. (There are undoubtedly many other methods that would achieve the same effect.)[14] Hence, aggregate L0 Hider behaviour is given by

$$p_{0H}(w) = w_L\left(\frac{1}{6}, \frac{5}{12}, \frac{5}{12}\right) + w_C\left(\frac{5}{12}, \frac{5}{12}, \frac{1}{6}\right) + (1 - w_L - w_C)\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right).$$

Assuming (as before) that $w_L > w_C$, an L0 Hider is most likely to hide the object in the 'plain' box, the white X one (2) and least likely to hide it in the Z box.

Next, both L1 Hiders and L1 Seekers with frame L will pick among the X boxes: the Hider because she believes the L0 Seeker is most likely to look in the Z box; the Seeker because she believes the L0 Hider is likely to pick an X box. L1 Hiders and Seekers with frame C will pick among the white boxes: the Hider because she believes the L0 Seeker is most likely to look in the off-white box; the Seeker because she believes the L0 Hider is likely to pick a white box.

L1 Hiders and Seekers with frame LC will both choose the 'least salient' box, the white X box (2): the Hider because she believes the L0 Seeker is least likely to look there; the Seeker because she believes the L0 Hider is more likely to hide there. Hence, for all frames, L1 Hiders and Seekers exhibit the same behaviour, and the aggregate L1 behaviour is given by

$$p_{1H}(w) = p_{1S}(w) = w_L\left(0, \frac{1}{2}, \frac{1}{2}\right) + w_C\left(\frac{1}{2}, \frac{1}{2}, 0\right) + (1 - w_L - w_C)(0, 1, 0).$$

Clearly, the modal behaviour of L1 types is to pick the least salient box (2). Thus, L1 behaviour is the same as Rubinstein *et al.*'s 'naive strategies' and is consistent with the modal choice in their experiment. However, it fails to explain why in their experiment Seekers tended more strongly than Hiders to choose box 2. Indeed, combined with L0 behaviour, there is already a bias toward Hiders choosing box 2 more often than Seekers. But we have yet to consider L2 types.

[13] For early humans, 'looking' problems were more generic and 'hiding' problems more strategic. Early humans looked for fruits and vegetables, water, shelter — non-strategic decision problems — while they hid from animal predators and human enemies.

[14] The selective pressure for hider-saliencies of the above kind, produced by these ecological circumstances, 'hard-wired' a simple form of level 1 reasoning. This explains why level 0 Hiders, who can do no explicit modelling of their opponents, may nevertheless behave in a way which reflects thier strategic situation.

Unlike L1 types, Hiders and Seekers of L2 type will differ in their behaviour. First, consider L2 Hiders. Given frame L, L0 and L1 Seekers have conflicting tendencies, so a L2 Hider's behaviour will depend on her belief, $\widehat{\alpha}_0$, about the proportion of L0 types in the Seeker population. Her probabilities for the Seeker's choice are

$$q_{2H}(L) = \widehat{\alpha}_0 \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right) + (1 - \widehat{\alpha}_0) \left(0, \frac{1}{2}, \frac{1}{2}\right).$$

If $\widehat{\alpha}_0 < 2/3$, then the Z box is believed to be least likely, and the X boxes are equally likely; hence, for these parameters we predict that L2 Hiders with frame L will choose the Z box.

Given frame C, an L2 Hider's probabilities for the Seeker's choice are

$$q_{2H}(C) = \widehat{\alpha}_0 \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}\right) + (1 - \widehat{\alpha}_0) \left(\frac{1}{2}, \frac{1}{2}, 0\right).$$

If $\widehat{\alpha}_0 < 2/3$, then the off-white X box is believed to be least likely; hence, for these parameters we predict that L2 Hiders with frame C will choose the box 3.

Given frame LC, since both L0 and L1 Seekers have conflicting tendencies, an L2 Hider's behaviour will depend on her belief, $\widehat{\alpha}_0$, about the proportion of L0 types in the Seeker population, and $(\widehat{w}_L, \widehat{w}_C)$, her beliefs about the probability that the Seeker has frame L or C respectively. Then, her probabilities for the Seeker's choice are

$$q_{2H}(\widehat{w} \mid LC) = \widehat{\alpha}_0 p_{0S}(\widehat{w}) + (1 - \widehat{\alpha}_0) p_{1S}(\widehat{w}).$$

Given $\widehat{w}_L > \widehat{w}_C$ and $\widehat{\alpha}_0 < 2/3$, these probabilities are ordered as $q_{2H}(2; \widehat{w} \mid LC) > q_{2H}(3; \widehat{w} \mid LC) > q_{2H}(1; \widehat{w} \mid LC)$. Therefore, for these parameters, L2 Hiders with frame LC will choose Z, the most salient box.

Combining the results for L2 Hiders, the most salient box 1 will be the modal choice when $w_L > w_C$, and the least salient box 2 will be avoided entirely.

Finally consider L2 Seekers. Given frame L, both L0 and L1 Hiders tend to avoid the Z box, but are equally likely to choose an X box. Therefore, a L2 Seeker with frame L will pick an X box. Given frame C, both L0 and L1 Hiders tend to hide the object in the off-white X box (3), so a L2 Seeker with frame C will choose box 3.

Given frame LC, since L0 and L1 Hiders have conflicting tendencies, an L2 Seeker's behaviour will depend on her $\widehat{\alpha}_0$ and $(\widehat{w}_L, \widehat{w}_C)$. Then, her probabilities for the Hider's choice are

$$q_{2S}(\widehat{w} \mid LC) = \widehat{\alpha}_0 p_{0H}(\widehat{w}) + (1 - \widehat{\alpha}_0) p_{1H}(\widehat{w}).$$

Given $\widehat{w}_L > \widehat{w}_C$, and $\widehat{\alpha}_0 < 2/3$, these probabilities are ordered as $q_{2S}(2; \widehat{w} \mid LC) > q_{2S}(3; \widehat{w} \mid LC) > q_{2S}(1; \widehat{w} \mid LC)$. Therefore, for these parameters, L2 Seekers with frame LC will choose the least salient box (2). Combining the results for L2 Seekers, the least salient box will be the modal choice when $w_L + 2w_C < 1$, which is quite reasonable for this Hide and Seek game.

Thus, we see that L2 Seekers are much more likely to choose the least salient box (2) than L2 Hiders. Therefore, if the population has more L2 types than L0 types, then this extended VFLNT would predict that the least salient box is the modal choice, and more strongly for Seekers than Hiders. This later prediction hinges on our introduction of the salience-avoiding L0 Hider type. The fact that this assumption leads to a model that explains the data better complements our initial defence of this assumption.

5 Testing VFLNT

The basic framing effects which appear in VFLNT are the probabilities of the frames $w(f)$ and the frame beliefs $\widehat{w}(f' \mid f)$. In VFT, the corresponding variables are $v(f)$ and $\widehat{v}(f' \mid f)$. Call these variables collectively the *framing parameters*.

To test VFLNT we suggest using a two-stage procedure. First, estimate the framing parameters from behaviour in certain ancillary choice tasks called 'Buridan tasks'; then, inserting these estimates in VFLNT, test the predictions of this model against behaviour in a variety of games. As a byproduct of these tests, we would obtain estimates of population proportions of players of different types, as in [SW95]. Since one type of player (the E type) obeys rationalistic VFT, these estimated type-proportions provide, among other things, a test of that model against boundedly-rational alternatives.

In a simple Buridan Task (BT) a subject is presented with an object set A and asked to choose any one to obtain a fixed reward. In the Prompted version (PBT) she is also prompted to think about the characteristics of the objects. The subject's opportunities to randomize are kept to a minimum. In a Buridan Estimation Task (BET), the subject is shown A and asked to estimate the percentages who make various choices in the BT (or PBT) with object set A.

We suggest that agents choose in a BT with set A much as L0 players do in a matching game with set A (recall, L0 players do not think strategically). We could therefore use the relative frequencies of choices in a prompted BT with object set A to estimate $w(*)$ from (1).

We also suggest that in a BET with set A subjects would respond in a manner that would reveal potential noticer bias. Using BT and BET data on set A and (3), it should be possible to obtain estimates of $\widehat{w}(* \mid f)$.

6 Conclusion

We have presented a Variable Frame Level-N model of behaviour in games that combines the fully rational Variable Frame Theory of [B93] and the boundedly rational Level-N Theory of [SW95]. In the process, we have demonstrated that the variable frame approach has important and significant effects independent of the rationality assumptions of VFT. For the most part VFLNT makes the same predictions as VFT in matching games. Most notably, VFLNT explains the Schelling competence. Instead of relying on extended rationality concepts, the Schelling competence in VFLNT arises from the unthinking tendency of level-0 types to choose a salient object, and then the boundedly rational deliberations of higher level types magnifies this initial tilt. Thus, the resolution of coordination indeterminacies offered by rational VFT is more a product of the variable frame apparatus than of the rationality assumptions.

We have also indicated how VFLNT can be extended to a non-coordination game (Hide and Seek), where it could offer new explanations of observed behaviour. Finally, believing that VFLNT has empirical content, we have suggested a strategy for testing it, which we are pursuing.

References

[B93] Michael Bacharach 1993, ‘Variable Universe Games’, in K. Binmore. A. Kirman and P. Tani (eds.), *Frontiers of Game Theory*, MIT Press.

[B96] Michael Bacharach 1996, ‘When There’s Nothing To Choose: A Model of Choice in Buridan Problems and an Experimental Design’, Economic Beliefs and Behaviour Discussion Paper, University of Kent.

[BB97] Michael Bacharach and Michele Bernasconi 1997, ‘The Variable Frame Theory of Focal Points: An Experimental Study’, *Games and Economic Behavior* **19**, 1-45.

[B82] D. E. Broadbent 1982, ‘Task Combination and Selective Intake of Information’, *Acta Psychologica* **50**, 253-290.

[EC66] G. Evans and C. Crumbaugh 1966, ‘Effects of Prisoner’s Dilemma Format on Cooperative Behaviour’, *Journal of Personality and Social Psychology* **3**, 486-488.

[G75] David Gauthier 1975, ‘Coordination’, *Dialogue* **14**, 195-221.

[G89] Margaret Gilbert 1989, ‘Rationality and Salience’, *Philosophical Studies* **57**, 61-77.

[H78] Jane Heal 1978, ‘Common Knowledge’, *Philosophical Quarterly* **28**, 116-31.

[J96] Maarten Janssen 1996, ‘Rationalizing Focal Points’, working paper, Dept. of Economics, Erasmus University.

[NS72] A. Newell and H. A. Simon 1972, *Human Problem Solving*, Prentice-Hall.

[RTH95] Ariel Rubinstein, Amos Tversky and Dana Heller 1995, ‘Naive Strategies in Competitive Games’, working paper, The Foerder Institute for Economic Research, Tel-Aviv University, 17 July 1995.

[M96] Judith Mehta 1996, *Bargaining Behaviour and the Social Foundations of Rational Choice*, Ph.D. dissertation, School of Social and Economic Studies, University of East Anglia.

[S60] Thomas Schelling 1960, *The Strategy of Conflict*, Harvard University Press.

[SW95] Dale Stahl and Wilson 1995, ‘On Players’ Models of Other Players: Theory and Experimental Evidence’, *Journal of Economic Behavior and Organization* **25**, 309-327.

[S95] Robert Sugden 1995, ‘A Theory of Focal Points’, *Economic Journal* **105**, 533-550.

[HBB90] John Van Huyck, Raymond Battalio, and Richard Beil 1990, 'Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure', *American Economic Review* **80**, 234-248.

[W1874] Wilhelm Wundt 1874, *Grundzuege der Physiologischen Psychologie* [*Principles of Physiological Psychology*], Springer.