Frustration and Anger in Games*

Pierpaolo Battigalli†  Martin Dufwenberg‡  Alec Smith§

January 4, 2015

Abstract

The economic consequences of anger may be important, concerning e.g. pricing, traffic safety, violence, and politics. Drawing on insights from psychology, we develop a formal approach to exploring how frustration and anger, via blame and aggression, shape interaction and outcomes in economic settings.

KEYWORDS: frustration, anger, blame, belief-dependent preferences, psychological games

---

*Preliminary version, missing one section, prepared for a seminar at UCSB on January 7, 2015. This paper modifies and extends Smith (2009). Pierpaolo Battigalli gratefully acknowledges financial support from ERC advanced grant 324219. We thank Paolo Leonetti and Marco Stenborg Pettersson for excellent research assistance and multiple seminar audiences for helpful comments.

†Bocconi University and IGIER.
‡Bocconi University, IGIER, University of Arizona, University of Gothenburg, CESifo.
§University of Arizona and Compass Lexecon.
1 Introduction

Anger may play a key role for shaping outcomes in economically important ways. Consider three cases:

Case 1: In 2006 US gas prices went up & up. Many folks were upset. Did this cause road rage, or people trading a truck for a Hyundai? Did gas stations or truck dealers go easier on price hikes, or offer rebates, anticipating potential adverse effects on sales that might otherwise materialize?

Case 2: When local football teams favored to win lose, the police get more reports of husbands assaulting wives (Card & Dahl 2011). Do unexpected losses spur thus vented frustration?

Case 3: Following Sovereign Debt Crises (2009-), some EU countries embarked on austerity programs. Was it because citizens lost benefits that some cities experienced riots?

Traffic safety, pricing, domestic violence, political landscapes, ... the themes seem important. However, in order to systematically assess relevance and consequences one needs theory connecting anger, decisions, and outcomes. In this paper we develop this.

Insights from psychology suggest ways that anger has strategic implications. The behavioral consequences of emotions are referred to as “action tendencies,” and that of anger is aggression. One may imagine that angry players are willing to forego material gains to punish others, or that a predisposition to behave aggressively when angered may benefit a player by serving as a credible threat, and so on. But while insights of this nature can be gleaned from psychologists’ writings, their analysis usually stops with the individual rather than going on to assess overall economic implications. We take the basic insights about anger that psychology has produced as input and inspiration for the theory we develop and apply.

1Baumeister & Bushman (2007, p. 66) e.g. say that “anger is an important and powerful cause of aggression” defined (p. 62) as “any behavior that is intended to harm another person who is motivated to avoid the harm.”

2The relevant literature is huge. A good point of entry, and source of insights and inspiration for us, is the recent International Handbook of Anger (Poteagal et al. 2010), which offers a cross-disciplinary perspective over 32 chapters reflecting “affective neuroscience, business administration, epidemiology, health science, linguistics, political science,
Economists traditionally paid scant attention to anger, but interest is on the rise and several recent studies inspire us. Most are empirical, indicative of hostile action occurring in economic situations, based on either archival or experimental data. A few of these studies present theory, typically with the purpose of explaining specific data patterns. Our approach is different. We do not start with data but with notions from psychology that we incorporate in general games. We are lead to models that differ substantially from the existing theory, though predictions may be similar in their specific settings.

Psychologists suggest that anger is typically anchored in frustration, which occurs when someone is unexpectedly denied something they care about. We assume that people are frustrated when they get less material rewards than they expected beforehand, and that they become hostile towards whomever they blame. However, there are several ways that blame may be assigned (cf. Alicke 2000) and we present three distinct approaches, captured by distinct utility functions. While players motivated by simple anger (SA) become generally hostile, those motivated by anger from blaming behavior (ABB) or by anger from blaming intentions (ABI) go after others more discriminately asking who caused, or who intended to cause, their dismay.

What are the overall implications when people interact? To provide answers, we develop a notion of polymorphic sequential equilibrium (PSE). Players are assumed to correctly anticipate how others behave on average, and the concept furthermore allows for different “types” of the same player to have different plans in equilibrium, which yields meaningful updating of players’ views of others’ intentions as various subgames are reached. This is crucial for a sensible treatment of how players consider intentionality as they blame others. We apply this solution concept to the aforementioned utility functions, explore properties, and compare predictions.

A player’s frustration depends on his beliefs about others’ choices. The blame a player attributes to another may depend on his beliefs about others’ choices or beliefs. For these reasons, all our models find their intellectual home in the framework of psychological game theory; see Geanakoplos et al. psychology, psychophysiology, and sociology” (p. 3, opening chapter). We take the non-occurrence of “economics” in the list as an indication our approach is original and needed.

6Psychologists often refer to this as “goal-blockage”; cf. p.3 of the (op.cit.) Handbook.
We develop most of our analysis within a two-period setting described in Section 2. Section 3 defines frustration. Section 4 develops our three key notions of psychological utility. Section 5 introduces the equilibrium concept and derives/highlight various results and insights. Section 6 generalizes the analysis to multistage games. Section 7 concludes.

2 Setup

Players engage in a two-stage interaction. Stage, or period $t \in \{1, 2\}$ is the time interval between dates $t-1$ and $t$. The set of active players and their feasible actions depend on the period and on previous choices. Players start with initial beliefs at date zero, and revise their beliefs conditioning on what they learn. We first describe the rules of interaction, or game form, and then we define initial and conditional beliefs.

2.1 Game form

We consider a finite two-stage game form describing the rules of interaction and the consequences of players’ actions. The set of players, possibly including passive individuals, is $I$. Letting $\emptyset$ denote the empty history (the root of the game), there is a finite set $I(\emptyset)$ of first movers; each $i \in I(\emptyset)$ picks an action $a^1_i$ in the finite feasible set $A_i(\emptyset)$. The finite set of active players in the second period, $I(a^1)$, depends on the first-period action profile $a^1 = (a^1_i)_{i \in I(\emptyset)}$, which becomes public information at the beginning of the second period. If $I(a^1) = \emptyset$, the game ends; otherwise, each player $i \in I(a^1)$ chooses an action $a^2_i$ in the finite feasible set $A_i(a^1)$; the resulting action profile is $a^2 = (a^2_i)_{i \in I(a^1)}$. In each period, active players move simultaneously. We let $A(\emptyset) = \times_{i \in I(\emptyset)} A_i(\emptyset)$ and $A(a^1) = \times_{i \in I(\emptyset)} A_i(a^1)$ (for each $a^1 \in A(\emptyset)$) denote the sets feasible action profiles. Similarly, for each player $i$, $A_{-i}(\emptyset) = \times_{j \in I(\emptyset) \setminus \{i\}} A_j(\emptyset)$ and $A_{-i}(a^1) = \times_{j \in I(\emptyset) \setminus \{i\}} A_j(a^1)$ denote the sets of feasible action profiles of the co-players.\footnote{By convention $A_j(h)$ is a singleton (a set containing only the pseudo-action "wait") whenever $j \notin I(h)$. Thus $|A_{-i}(h)| = 1$ if $I(h) = \{i\}$. Whenever a set $Y$ is a singleton, we identify the Cartesian product $X \times Y$ with $X$ as the two are isomorphic.}

The root $\emptyset$ and the feasible histories $a^1 \in A(\emptyset)$, $(a^1, a^2) \in A(\emptyset) \times A(a^1)$ are the nodes of the game tree. We let $Z = Z_1 \cup Z_2$ denote the set of...
terminal histories/nodes of the game tree, where $Z_t$ is the set of terminal histories of length $t$.\(^8\) Therefore, the set of non-terminal, or partial histories is $H = \{\emptyset\} \cup A(a^1) \setminus Z_1$. For each $h \in H$, $Z(h)$ denotes the set of terminal successors of $h$.\(^9\) If the set of active players $I(h)$ is a singleton for each $h \in H$, that is, if players never move simultaneously, then the game has perfect information.

We assume that the consequences of players’ actions are monetary gains (or losses), determined by the profile of material payoff functions $(\pi_i : Z \to \mathbb{R})_{i \in I}$. This completes the description of the game form, if there are no chance moves. If the game contains chance moves, we augment the player set with a dummy player $c$ (with $c \notin I$), who selects a feasible action at random. Thus, we consider an augmented player set $I_c = I \cup \{c\}$, and the sets of first movers and second movers may include $c$ as well: $I(\emptyset), I(a^1) \subseteq I_c$, $a^1 = (a^1)_i \in A(\emptyset)$, $a^2 = (a^2)_i \in A(a^1)$. If the chance player is active at history $h \in \{\emptyset\} \cup A(\emptyset)$, the chance move is described by a probability density function (pdf) $\sigma_c(h) \in \Delta(A_c(h))$. Table 1 summarizes:

<table>
<thead>
<tr>
<th>Notation</th>
<th>Terminology</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i \in I$</td>
<td>players</td>
</tr>
<tr>
<td>$c, I_c = I \cup {c}$</td>
<td>chance, set of players including chance</td>
</tr>
<tr>
<td>$t \in {1, 2}$</td>
<td>stages, or periods</td>
</tr>
<tr>
<td>$a^t_i$</td>
<td>action of $i$ in stage $t$</td>
</tr>
<tr>
<td>$a^t$ ($a^t_{-i}$)</td>
<td>action profile (of others) in stage $t$</td>
</tr>
<tr>
<td>$h \in H$</td>
<td>non-terminal, or partial histories</td>
</tr>
<tr>
<td>$I(h) \subseteq I_c$</td>
<td>set of active players at $h$</td>
</tr>
<tr>
<td>$A_i(h), A(h), A_{-i}(h)$</td>
<td>set of actions and action profiles at $h$</td>
</tr>
<tr>
<td>$\sigma_c(a^t_i</td>
<td>h)$</td>
</tr>
<tr>
<td>$z \in Z = Z_1 \cup Z_2$</td>
<td>terminal histories (of length 1 and 2)</td>
</tr>
<tr>
<td>$Z(h)$</td>
<td>terminal successors of $h$</td>
</tr>
<tr>
<td>$\pi_i : Z \to \mathbb{R}$</td>
<td>monetary payoff function of $i \in I$</td>
</tr>
</tbody>
</table>

Table 1. Elements of the two-stage game form.

The following example illustrates our notation.

\(^8\) $Z_1 := \{a^1 : a^1 \in \times_{i \in I(\emptyset)} A_i(\emptyset), I(a^1) = \emptyset\}$,

$Z_2 := \{(a^1, a^2) : a^1 \in \times_{i \in I(\emptyset)} A_i(\emptyset), I(a^1) \neq \emptyset, a^2 \in \times_{i \in I(a^1)} A_i(a^1)\}$

\(^9\) That is, $Z(\emptyset) = Z$ and $Z(\bar{a}^1) = \{(a^1, a^2) : a^1 = \bar{a}^1, a^2 \in A(\bar{a}^1)\}$. 

Example 1 (Asymmetric Punishment) Ann, Bob and Penny the punisher play the following game form, where Ann and Bob move simultaneously in the first stage and Penny may move in the second (profiles of actions and of monetary payoffs are listed in alphabetical order):

Using our notation, we have:

\[ H = \{ \emptyset, (D, L) \} , \]
\[ Z = \{(U, L), (U, R), (D, L), ((D, L), N), ((D, L), P)\} , \]
\[ I(\emptyset) = \{a, b\}, I((D, L)) = \{p\} \]
\[ A_a(\emptyset) = \{U, D\}, A_b(\emptyset) = \{L, R\}, A_p((D, L)) = \{N, P\}. \]

Penny – if active – can decrease \( \pi_b \), but not \( \pi_a \). Furthermore, punishing Bob is costly for Penny and rewards Ann.

2.2 Beliefs

It is conceptually useful to distinguish the following three aspects of a player’s beliefs: beliefs about co-players’ actions, beliefs about co-players’ beliefs, and a player’s plan, which we represent as beliefs about his own choices. Beliefs are formed conditional on each history. Let us abstractly denote by \( \Delta_{-i} \) the
space of co-players’ beliefs. Player $i$’s beliefs can be compactly described as conditional probability measures over paths and beliefs of others, that is, over $Z \times \Delta_{-i}$. Events, from $i$’s point of view, are subsets of $Z \times \Delta_{-i}$. Events about behavior have the form $Y \times \Delta_{-i}$, with $Y \subseteq Z$; events about beliefs have the form $Z \times E_{\Delta_{-i}}$, with $E_{\Delta_{-i}} \subseteq \Delta_{-i}$.$^{10}$

**Personal histories** To model how $i$ determines the subjective value of each feasible action at each history where he is active, we add to the commonly observed histories $h \in H$ also personal histories of the form $(h;a_i)$, with $i \in I(h)$, $a_i \in A_i(h)$. In a game with perfect information, $(h,a_i) \in H \cup Z$. But if there are simultaneous moves at $h$, then $(h,a_i)$ is not a history in the standard sense. As soon as $i$ irreversibly chooses action $a_i$, he observes $(h,a_i)$, and determines the value of $a_i$ using his beliefs conditional on this event. We denote by $H_i$ the set of histories of $i$ – the standard and personal ones – and by $Z((h;a_i))$ the set of terminal successors of personal history $(h,a_i)$. The standard precedence relation $\prec$ for histories in $H \cup Z$ is extended to $H_i$ in the obvious way: for all $h \in H$, $i \in I(h)$, and $a_i \in A_i(h)$, it holds that $h \prec (h,a_i)$ and $(h,a_i) \prec (h,(a_i,a_{-i}))$ if $i$ is not the only active player at $h$.

**Conditional probability systems** Player $i$’s system of beliefs $\mu_i$ is an array of conditional beliefs indexed by histories in $H_i$: $\mu_i = (\mu_i(\cdot|[h_i]))_{h_i \in H_i}$, where $\mu_i(\cdot|[h_i])$ is a probability measure concentrated on event about behavior $[h_i] = Z(h_i) \times \Delta_{-i}$ for all $h_i \in H_i$. We use obvious abbreviations like

$$\mu_i(h_i'|h_i) = \mu_i([h_i']|[h_i])$$

whenever this causes no confusion. More generally, we suppress parentheses when this does not compromise understanding.

The first-order belief system of $i$ gives the probabilities of terminal histories and of action profiles conditional on each history:

$$\alpha_i(z|h_i) = \mu_i(z|h_i) \quad \alpha_i(a|h) = \alpha_i((h,a)|h) \quad (1)$$

$^{10}$We assume that $\Delta_{-i}$ is a compact metrizable space, which is justified by the construction of hierarchical belief spaces given below. Events are Borel measurable subsets of $Z \times \Delta_{-i}$. We do not specify the terminal beliefs of $i$ about the beliefs of others, because they are not relevant for the models in this paper.

$^{11}H_i = H \cup \{(h,a_i) : h \in H, i \in I(h), a_i \in A_i(h)\}$;

$$Z(h,a_i) = \bigcup_{a_{-i} \in A_{-i}(h)} Z(h,(a_i,a_{-i})).$$
for all \( z \in Z \), \( h_i \in H_i \), \( h \in H \) and \( a \in A(h) \).

\( \mu_i \) must satisfy some natural properties. First of all, the rules of conditional probabilities must hold whenever possible: if \( h_i \prec h'_i \) then

\[
\mu_i(h'_i|h_i) > 0 \Rightarrow \mu_i(E|h'_i) = \frac{\mu_i(E \cap [h'_i]|h_i)}{\mu_i(h'_i|h_i)} \tag{2}
\]

for all \( h_i, h'_i \in H_i \) and every event \( E \subseteq Z \times \Delta_{-i} \). (1)-(2) imply

\[
\alpha_i(a^1, a^2|\emptyset) = \alpha_i(a^2|a^1) \alpha_i(a^1|\emptyset).
\]

Second, \( i \) realizes that his choice cannot influence simultaneous choices and beliefs of co-players, so \( i \)'s beliefs satisfy a causal independence property:

\[
\mu_i([h, (a_i, a_{-i})] \cap E_{-i}(h, a_i)) = \mu_i([h, (a'_i, a_{-i})] \cap E_{-i}(h, a'_i)) \tag{3}
\]

for every \( h \in H \), \( a_i, a'_i \in A_i(h) \), \( a_{-i} \in A_{-i}(h) \), and \( E_{-i} = Z \times E_{\Delta_{-i}} \).

(1)-(3) imply

\[
\alpha_i(a_i, a_{-i}|h) = \alpha_{i,i}(a_i|h) \times \alpha_{i,-i}(a_{-i}|h),
\]

where \( \alpha_{i,i}(|h) \) and \( \alpha_{i,-i}(|h) \) are marginals of \( \alpha_i(|i) \) on \( A_i(h) \) and \( A_{-i}(h) \).

Note that the array of conditional probabilities \( \alpha_i = (\alpha_{i,i}(|i))_{h \in H} \in \times_{h \in H} \Delta_i(A_i(h)) \) is – technically speaking – a behavioral strategy, and we interpret it as the plan of \( i \). The reason is that the result of \( i \)'s contingent planning is precisely a system of conditional beliefs about what action he would take at each history. If there is only one co-player, also \( \alpha_{i,-i} \in \times_{h \in H} \Delta_i((A_{-i}(h)) \) formally corresponds to a behavioral strategy. With multiple co-players, \( \alpha_{i,-i} \) corresponds instead to a “correlated behavioral strategy.” Whatever the case, \( \alpha_{i,-i} \) gives the conditional beliefs of \( i \) about the behavior of others, and these beliefs may not coincide with the plans of others. We emphasize that the plan of a player is not an actual choice; actions on the path of play are the only actual choices.

A belief system \( \mu_i \) satisfying (2)-(3) is a conditional probability system, or CPS. The set of such CPSs is denoted \( \Delta^{H_i}(Z \times \Delta_{-i}) \), a subset of \( [\Delta(Z \times \Delta_{-i})]^{H_i} \). Whenever this causes no confusion, we write initial beliefs omitting the empty history, as in \( \mu_i(E) = \mu_i(E|\emptyset) \), or \( \alpha_i(a) = \alpha_i(a|\emptyset) \).
Hierarchical beliefs Of course, (2)-(3) can be directly stated for the system of first-order beliefs \( \alpha_i = (\alpha_i(\cdot|h))_{h \in H_i} \), that is, the conditional beliefs about paths. The set of first-order CPSs, \( \Delta_1^i = \Delta_j^H(Z) \), is a compact metrizable space. The set of second-order CPSs, \( \Delta_2^i = \Delta_j^H(Z \times \Delta_1^i) \) where \( \Delta_1^i = \times_{j \in I \setminus \{i\}} \Delta_j^1 \), is compact and metrizable as well. Higher-order belief spaces can be defined by recursion, but we do not need them in our analysis. Note that the first-order CPS \( \alpha_i \in \Delta_1^i = \Delta_j^H(Z) \) has to be derived from the second-order CPS \( \beta_i \in \Delta_2^i = \Delta_j^H(Z \times \Delta_1^i) \), otherwise \( i \)'s second-order hierarchy \( (\alpha_i, \beta_i) \) would be incoherent. Indeed, it can be checked that starting from \( \beta_i \in \Delta_2^i \) and letting

\[
\alpha_i(Y|h) = \beta_i(Y \times \Delta_1^i|h)
\]

for all \( h \in H_i \) and \( Y \subseteq Z \), we obtain an array of conditional probabilities satisfying (1)-(3), that is, an element of \( \Delta_1^i \). Whenever we write in a formula beliefs of different orders for the same player, we assume that first-order beliefs are derived from second-order beliefs.

Conditional expectations Let \( \psi_i \) be any real-valued measurable function of variables that player \( i \) does not know, e.g., the terminal history and the co-players’ beliefs of order \( \ell \). Then \( i \) can compute the expected value of \( \psi_i \) conditional on any common or personal history \( h_i \in H_i \) by means of his \( m \)-order CPS \( \mu_i^m \), provided that \( m > \ell \) (in particular, we consider \( m = 2, \ell = 1 \)). This expected value is denoted by \( \mathbb{E}[\psi_i|h_i; \mu_i^m] \). The first-order CPS \( \alpha_i \) gives the conditional expected material payoffs:

\[
\mathbb{E}[\pi_i|h; \alpha_i] = \sum_{z \in Z(h)} \alpha_i(z|h)\pi_i(z),
\]

\[
\mathbb{E}[\pi_i|(h, a_i); \alpha_i] = \sum_{z \in Z(h,a_i)} \alpha_i(z|h,a_i)\pi_i(z)
\]

for all \( h \in H, a_i \in A_i(h) \). For initial beliefs, we omit \( h = \emptyset \) from such expressions; in particular, the initially expected payoff is \( \mathbb{E}[\pi_i; \alpha_i] \).

---

12 This holds for higher-order beliefs in general; following a proof by Battigalli & Siniscalchi (1999) with a minor modification to take independence into account.
3 Frustration

We will present several models of how frustrated players attribute blame and go after others, but keep our account of frustration constant. Here is the key definition: $i$’s frustration in stage 2, given $a^1$ is given by

$$F_i(a^1; \alpha_i) = \left[ \mathbb{E}[\pi_i; \alpha_i] - \max_{a^2_i \in A_i(a^1)} \mathbb{E}[\pi_i|(a^1, a^2_i); \alpha_i] \right]^+,$$

where $[x]^+ = \max\{x, 0\}$. $i$’s frustration in stage 2 is given by the gap, if positive, between his initially expected payoff and the currently best expected payoff he believes he can obtain. Three comments are warranted:

1. Diminished expectations – $\mathbb{E}[\pi_i|a^1; \alpha_i] < \mathbb{E}[\pi_i; \alpha_i]$ – is only a necessary condition for frustration. For $i$ to be frustrated it must also be that $i$ cannot close the gap. Had we alternatively modeled frustration as equal to actual diminished expectations ($\mathbb{E}[\pi_i; \alpha_i] - \mathbb{E}[\pi_i|a^1; \alpha_i]$) this would have counterintuitive implications.

2. $F_i(a^1; \alpha_i)$ expresses stage 2 frustration. One could define frustration at the root, or at end nodes, but neither would matter for our purposes. At the root nothing happened so frustration equals zero. Frustration is possible at the end nodes, but can’t influence subsequent choices as the game is over. One might have thought that anticipated frustration at end nodes matters to earlier decisions; however, to simplify the analysis the assumptions we make (below) rule this out. Players care in advance only about future frustrations of others only insofar as they move behavior.

3. The psychological evidence, cited in the introduction, says that a player becomes frustrated when his goals are unexpectedly thwarted. $F_i(a^1; \alpha_i)$ captures one aspect, concerning own material rewards. Cases 1-3 of the introduction indicate the broad applied potential. Nevertheless, our focus is restrictive, in that one may imagine sources of frustration other than expected material payoff. To see this, consider two more cases:

Case 4: In 2007 Apple launched its iPhone at $499$. Two months later they introduced a new version at $399$, re-priced the old model at $299$, and caused outrage among early adopters. Apple paid back the difference. Did this help long run profit?
Case 5: The 2008 TARP bank bail-out infuriated some US voters. Did this ignite the Tea Party/Occupy-WS movements?

Cases 4-5 make sense but our definition does not address them (which is why we used other cases in the introduction). In case 4 an early adapter is frustrated because he regrets he already bought, not because new information implies that his expected rewards drop. In case 5, an activist may be materially unaffected personally, yet frustrated because of unexpected perceived unfairness. Our analysis addresses many subtle considerations, yet some meaningful ways to get frustrated are left for future research.

To illustrate the definition of $F_i(a^i; \alpha_i)$, return to the game of Example 1

Example 2 Suppose that, in the game form of Figure A, Penny initially expects to get $82, i.e., \( \alpha_p((U,L)|\emptyset) + \alpha_p((D,R)|\emptyset) = 1 \) and \( \mathbb{E}[\pi_p; \alpha_p] = 2 \). Penny’s frustration after $a^1 = (D,L)$ is

\[
F_p((D,L); \alpha_i) = [\mathbb{E}[\pi_p; \alpha_p] - \max\{\pi_p((D,L), N), \pi_p((D,L), P)\}]^+ = 2 - 1 = 1.
\]

Penny’s frustration is independent of her plan, because she is initially certain she will not move. Suppose instead that \( \alpha_p((U,L)|\emptyset) = \alpha_p((D,L)|\emptyset) = \frac{1}{2} \). Then

\[
F_p((D,L); \alpha_i) = \frac{1}{2} \times 2 + \frac{1}{2} \alpha_p(N|(D,L)) \times 1 - 1 = \frac{1}{2} \alpha_p(N|(D,L)).
\]

Penny’s frustration is now highest when she plans not to punish Bob. Her frustration is independent of her actual choice though; Penny’s frustration equals \( \frac{1}{2} \alpha_p(N|(D,L)) \) independently of whether she ultimately chooses $N$ or $P$.

4 Anger

A player’s preferences over actions at a given node – his action tendencies – depend on expected material payoffs and frustration. A frustrated player tends to hurt others, if this is not too costly (cf. Dollard et al. 1939, Averill 1983, Berkowitz 1989). We consider different versions of this frustration/aggression hypothesis related to different levels of cognitive appraisal.
In general, player $i$ moving at history $h$ chooses action $a_i$ in order to maximize the expected value of a belief-dependent “decision utility” of the form

$$u_i(h, a_i; \mu^m_i) = \mathbb{E}[\pi_i(h, a_i ; \alpha_i) - \theta_i \sum_{j \neq i} B_{ij}(h; \mu^m_i) \mathbb{E}[\pi_j(h, a_i ; \alpha_i)]],$$

where $B_{ij}(h; \mu^m_i) \geq 0$ measures how much of $i$’s frustration is blamed on co-player $j$ (and hence the tendency to hurt $j$), $\alpha_i$ is the first-order CPS derived from $m$-order belief $\mu^m_i$, and $\theta_i$ is a sensitivity parameter. We assume that $B_{ij}(h; \mu^m_i)$ is positive only if frustration is positive:

$$B_{ij}(a^1; \mu^m_i) \leq F_i(a^1; \mu^m_i).$$

Therefore, the decision utility of a first-mover coincides with expected material payoff, because there cannot be any frustration in the first stage:

$$u_i(\emptyset, a_i; \mu^m_i) = \mathbb{E}[\pi_i|a_i; \alpha_i].$$

When $i$ is the only active player at $h = a^1$, he determines the terminal history with his choice $a_i = a^2$, and decision utility has the form

$$u_i(a^1, a_i; \mu^m_i) = \pi_i(a^1, a_i) - \theta_i \sum_{j \neq i} B_{ij}(h; \mu^m_i) \pi_j(a^1, a_i).$$

We now proceed to consider three specific functional forms that capture different notions of blame.

### 4.1 Simple Anger (SA)

Our most rudimentary hypothesis, simple anger (SA), is that $i$’s tendency to hurt others is proportional to $i$’s frustration, unmodulated by cognitive appraisal:

$$u_i^{SA}(a^1, a_i; \alpha_i) = \mathbb{E}[\pi_i|a^1, a_i] - \theta_i^{SA} \sum_{j \neq i} F_i(a^1, a_i) \mathbb{E}[\pi_j|a^1, a_i; \alpha_i].$$

**Example 3** (Ultimatum Minigame) Ann and Bob play a simple bargaining game: Ann can make a fair offer, which is automatically accepted, or a greedy
offer which Bob can either accept or reject.

\[ F_b(g; \alpha_b) = [(2(1 - \alpha_b(g)) + \alpha_b(g)b(y)) - 1]^+. \]

Therefore

\[ u_{b}^{SA}(g, n; \alpha_i) - u_{b}^{SA}(g, y; \alpha_i) = 3\theta_{b}^{SA} [(2(1 - \alpha_b(g)) + \alpha_b(g)b(y)) - 1]^+ - 1. \]

For Bob to be frustrated he must not expect the greedy offer with certainty. If he is frustrated, the less he expects the greedy offer, and – interestingly – the less he plans to reject it, the more prone he is to reject once the greedy offer materializes. The more resigned Bob is to getting a low payoff, the less frustrated and prone to aggression he will be when receiving the low-ball offer.

### 4.2 Anger from blaming behavior (ABB)

Action tendencies may depend on a player’s cognitive appraisal how to blame others. When a frustrated player \(i\) blames co-players for their behavior, he looks only at the actions chosen in stage 1, without considering intentions, that is, without considering others’ plans and beliefs about others. \(i\)’s tendency to hurt \(j\) is determined by a continuous blame function \(B_{ij}(a^1; \alpha_i)\) that depends only on first-order belief \(\alpha_i\) such that

\[ B_{ij}(a^1; \alpha_i) = \begin{cases} 0, & \text{if } j \notin I(\emptyset), \\ F_i(a^1; \alpha_i), & \text{if } \{j\} = I(\emptyset). \end{cases} \]

(5) says that if \(j\) is not active in the first stage, he cannot be blamed for \(i\)'s frustration and if instead \(j\) is the only active player he is fully blamed.
We consider below specific functional forms for $B_{ij}(a^1;\alpha_i)$ that satisfy (4)-(5). With this, the decision utility with anger from blaming behavior (ABB) is

$$u_i^{ABB}(a^1, a_i; \alpha_i) = \mathbb{E}[\pi_i(a^1, a_i; \alpha_i)] - \theta_i^{ABB} \sum_{j \neq i} B_{ij}(a^1; \alpha_i) \mathbb{E}[\pi_j(a^1, a_i; \alpha_i)].$$

The following example illustrates the difference between SA and ABB:

**Example 4** (Inspired by Frijda, 1993) Andy the handyman uses a hammer. His apprentice, Bob, has no payoﬀ-relevant action. In a bad day (determined by chance) Andy hammers his thumb and can then either take it out on Bob or not. If he does, he further disrupts production. See Figure C.

![Figure C. Hammering one’s thumb](image)

Assuming $\alpha_a(B) = \varepsilon < \frac{1}{2}$, the extent of Andy’s frustration in a bad day is

$$F_a(B; \alpha_a) = 2(1 - \varepsilon) + \varepsilon \alpha_a(N|B) - 1 > 0.$$  

With SA and $\theta_a^{SA}$ sufficiently high, Andy takes it out on Bob. But, since Bob is passive, Andy chooses $N$ regardless of $\theta_a^{SA}$.

SA and ABB yield the same behavior in the Ultimatum Minigame and similar game forms. Say that a game form is a leader-followers game if there is only one active player in the first stage, who does not move in stage two: $I(\emptyset) = \{j\}$ and $I(\emptyset) \cap I(a^1) = \emptyset$ for some $j \in I$ and every $a^1$. (5) implies:

**Remark 1** In leader-followers games, SA and ABB coincide, that is, $u_i^{SA} = u_i^{ABB}$ provided that $\theta_i^{SA} = \theta_i^{ABB}$.

Next, we contrast two specific functional forms for ABB.
**Could-have-been blame** When frustrated after action profile $a^1$, player $i$ considers, for each $j$, what he would have obtained at most, in expectation, had $j$ chosen differently:

$$\max_{a'_j \in A_j(\sigma)} \mathbb{E}[\pi_i|(a^1_{-j}, a'_j); \alpha_i].$$

If this could-have-been payoff is more than what $i$ currently expects ($\mathbb{E}[\pi_i|a^1; \alpha_i]$), then $i$ blames $j$, up to $i$'s frustration (so (4) holds):

$$B_{ij}(a^1; \alpha_i) = \min \left\{ \left[ \max_{a'_j \in A_j(\sigma)} \mathbb{E}[\pi_i|(a^1_{-j}, a'_j); \alpha_i] - \mathbb{E}[\pi_i|a^1; \alpha_i] \right]^+, F_i(a^1; \alpha_i) \right\},$$

(6)

If $j$ is not active in the first stage, then

$$\max_{a'_j \in A_j(\sigma)} \mathbb{E}[\pi_i|(a^1_{-j}, a'_j); \alpha_i] = \mathbb{E}[\pi_i|a^1; \alpha_i],$$

so

$$B_{ij}(a^1; \alpha_i) = \min \{0, F_i(a^1; \alpha_i)\} = 0.$$  

If instead $j$ is the only active player, then $B_{ij}(a^1; \alpha_i) = F_i(a^1; \alpha_i)$ since

$$\left[ \max_{a'_j \in A_j(\sigma)} \mathbb{E}[\pi_i|(a^1_{-j}, a'_j); \alpha_i] - \mathbb{E}[\pi_i|a^1; \alpha_i] \right]^+ \geq \left[ \mathbb{E}[\pi_i; \alpha_i] - \mathbb{E}[\pi_i|a^1; \alpha_i] \right]^+ \geq \left[ \mathbb{E}[\pi_i; \alpha_i] - \max_{a_i^2 \in A_i(a^1)} \mathbb{E}[\pi_i|(a^1, a_i^2); \alpha_i] \right]^+ = F_i(a^1; \alpha_i)$$

Therefore, function (6) satisfies (5).

**Example 5** Consider Penny at $a^1 = (D, L)$ in Figure A. For each $j \in \{a, b\}$, Penny’s could-have-been payoff is $2 \geq \mathbb{E}[\pi_p; \alpha_p]$, her expected payoff is $\mathbb{E}[\pi_p|(D, L); \alpha_p] \leq 1$, and her frustration is $\left[ \mathbb{E}[\pi_p; \alpha_p] - 1 \right]^+$. Therefore

$$B_{pa}((D, L); \alpha_p) = B_{pb}((D, L); \alpha_p) = \min \left\{ [2 - \mathbb{E}[\pi_p|(D, L); \alpha_p]^+, \left[ \mathbb{E}[\pi_p; \alpha_p] - 1 \right]^+] = \left[ \mathbb{E}[\pi_p; \alpha_p] - 1 \right]^+, \right\},$$

that is, both Ann and Bob are fully blamed by Penny for her frustration at $(D, L)$. 

15
Blaming unexpected deviations When frustrated after \(a^1\), \(i\) assesses, for each \(j\), how much he would have obtained had \(j\) behaved as expected:

\[
\sum_{a'_j \in A_j(\varnothing)} \alpha_{ij}(a'_j)\mathbb{E} \left[ \pi_i|(a_{-j}^1, a'_j); \alpha_i \right],
\]

where \(\alpha_{ij}(a'_j)\) is the marginal probability of action \(a'_j\) according to \(i\)'s belief \(\alpha_i\). With this, the blame formula is

\[
B_{ij}(a^1; \alpha_i) = \min \left\{ \left[ \sum_{a'_j \in A_j(\varnothing)} \alpha_{ij}(a'_j)\mathbb{E} \left[ \pi_i|(a_{-j}^1, a'_j); \alpha_i \right] - \mathbb{E}[\pi_i|a^1; \alpha_i] \right]^+, F_i(a^1; \alpha_i) \right\}. \tag{7}
\]

If \(j\) is not active in the first stage, we get

\[
B_{ij}(a^1; \alpha_i) = \min \left\{ \left[ \mathbb{E}[\pi_i|a^1; \alpha_i] - \mathbb{E}[\pi_i|a^1; \alpha_i] \right]^+, F_i(a^1; \alpha_i) \right\} = 0;
\]

\(j\) cannot have deviated and cannot be blamed. If, instead, \(j\) is the only active player in the first stage, then

\[
\sum_{a'_j \in A_j(\varnothing)} \alpha_{ij}(a'_j)\mathbb{E} \left[ \pi_i|(a_{-j}^1, a'_j); \alpha_i \right] = \sum_{a' \in A(\varnothing)} \alpha_i(a')\mathbb{E} \left[ \pi_i|a'; \alpha_i \right] = \mathbb{E} \left[ \pi_i|\alpha_i \right],
\]

and (7) yields

\[
B_{ij}(a^1; \alpha_i) = \min \left\{ \left[ \mathbb{E}[\pi_i|a^1; \alpha_i] - \mathbb{E}[\pi_i|a^1; \alpha_i] \right]^+, F_i(a^1; \alpha_i) \right\} = F_i(a^1; \alpha_i).
\]

Therefore, like blame function (6), also (7) satisfies (5).

If \(a_j^1\) is what \(i\) expected \(j\) to do in the first stage \((\alpha_{ij}(a_j^1) = 1)\) then

\[
B_{ij}(a^1; \alpha_i) = \min \left\{ \left[ \mathbb{E}[\pi_i|a^1; \alpha_i] - \mathbb{E}[\pi_i|a^1; \alpha_i] \right]^+, F_i(a^1; \alpha_i) \right\} = 0.
\]

In other words, \(j\) did not deviate from what \(i\) expected and \(j\) is not blamed by \(i\). This is different from “could-have-been” blame (6).

**Example 6** Suppose that, in Figure A, Penny is initially certain of \((U, L)\), so \(\alpha_p(U, L) = 1\) and \(\mathbb{E}[\pi_p|\alpha_p] = 2\). Upon observing \((D, L)\) her frustration is
\(F_p((D, L); \alpha_p) = [\mathbb{E}[\pi_p; \alpha_p] - 1]^+ = 1\). Using (7), at \(a^1 = (D, L)\), Penny fully blames Ann, who deviated from \(U\) to \(D\). Using that
\[
\sum_{a'_a \in A_a(\varnothing)} \alpha_p(a'_a) \mathbb{E} \left[ \pi_p | (a^1_{-a}, a'_a); \alpha_p \right] = \pi_p(U, L) = 2
\]
we get that Penny’s blame of Ann equals Penny’s frustration
\[
B_p((D, L); \alpha_p) = \min \left\{ \left[ 2 - \mathbb{E}[\pi_p|a^1; \alpha_p] \right]^+, 1 \right\} = 1.
\]

On the other hand, Penny does not blame Bob, who played \(L\) as expected. To verify this, note that when frustrated after \((D, L)\) Penny assesses how much she would have obtained had Bob behaved as expected:
\[
\sum_{a'_b \in A_b(\varnothing)} \alpha_p(a'_b) \mathbb{E} \left[ \pi_p | (a^1_{-b}, a'_b); \alpha_p \right] = \mathbb{E}[\pi_p|(D, L); \alpha_p]
\]
and
\[
B_p((D, L); \alpha_p) = \min \left\{ \left[ \mathbb{E}[\pi_p|(D, L); \alpha_p] - \mathbb{E}[\pi_p|(D, L); \alpha_p] \right]^+, 1 \right\} = 0,
\]
in contrast to could-have-been blame (5) under which, as we saw, Penny fully blames Bob (Example 5).

Blaming unexpected deviations and could-have-been blame both credit the full frustration on the first-mover of a leader-followers game, because they both satisfy (5) (see Remark 1).

### 4.3 Anger from blaming intentions (ABI)

A player \(i\) prone to anger from blaming intentions (ABI) asks himself, for each co-player \(j\), whether \(j\) intended to give him a low expected payoff. Since such intention depends on \(j\)’s first-order beliefs \(\alpha_j\) (which include \(j\)’s plan, \(\alpha_{j,j}\)), how much \(i\) blames \(j\) depends on \(i\)’s second-order beliefs \(\beta_i\), and the decision utility function has the form
\[
u^{\text{ABI}}_i(h, a_i; \beta_i) = \mathbb{E} \left[ \pi_i | (h, a_i); \alpha_i \right] - \theta^{\text{ABI}}_i \sum_{j \neq i} B_{ij} (h; \beta_i) \mathbb{E} \left[ \pi_j | (h, a_i); \alpha_i \right],
\]
where \(\alpha_i\) is derived from \(\beta_i\).
The maximum payoff that \( j \), initially, can expect to give to \( i \) is
\[
\max_{a_j \in A_j(\varnothing)} \sum_{a_{-j} \in A_{-j}(\varnothing)} \alpha_{j,-j}(a_{-j}) \mathbb{E} \left[ \pi_i \left| \left( a_j, a_{-j} \right) ; \alpha_j \right. \right].
\]

Note that
\[
\max_{a_j \in A_j(\varnothing)} \sum_{a_{-j} \in A_{-j}(\varnothing)} \alpha_{j,-j}(a_{-j}) \mathbb{E} \left[ \pi_i \left| \left( a_j, a_{-j} \right) ; \alpha_j \right. \right] \\
\geq \sum_{a_j \in A(j)} \alpha_j(a^1) \mathbb{E} \left[ \pi_i | a^1 ; \alpha_j \right] = \mathbb{E} \left[ \pi_i | \alpha_j \right],
\]
where the inequality holds by definition, and the equality is implied by the chain rule (2). Note also that \( \alpha_j(\cdot | a^1) \) is kept fixed under the maximization; we focus on what \( j \) considers he could achieve at the root, taking the view that he cannot control \( a_{-j}^2 \) but predicts how he will choose in stage 2. We assume that \( i \)'s blame on \( j \) at \( a^1 \) equals \( i \)'s expectation, conditional on \( a^1 \), of the difference between the maximum payoff that \( j \) can expect to give to \( i \) and what \( j \) plans/expects to give to \( i \), capped by \( i \)'s frustration:
\[
\mathbb{B}_{ij}(a^1; \beta_i) = \\
\min \left\{ F_i(a^1; \alpha_i), \int \left( \max_{a_{-j}} \sum_{a_{-j}} \alpha_{j,-j}(a_{-j}) \mathbb{E} \left[ \pi_i \left| \left( a_j, a_{-j} \right) ; \alpha_j \right. \right] - \mathbb{E} \left[ \pi_i | \alpha_j \right] \right) \beta_i(d \alpha_j | a^1) \right\},
\]
which is non-negative as per the previously highlighted inequality. Now, \( i \)'s decision utility after the first-stage history \( h = a^1 \) is
\[
u^{ABI}_i(h, a_i; \beta_i) = \mathbb{E} \left[ \pi_i \left| (h, a_i) ; \alpha_i \right. \right] - \theta^{ABI}_i \sum_{j \neq i} \mathbb{B}_{ij}(a^1; \beta_i) \mathbb{E} \left[ \pi_j \left| (a^1, a_i) ; \alpha_i \right. \right],
\]
where \(-\) in both equations \(-\alpha_i \) is derived from \( \beta_i \).

**Example 7** Consider the Ultimatum Minigame form of Figure B. The maximum payoff Ann can give to Bob is 2, independently of \( \alpha_a \). Suppose that Bob, upon observing the greedy offer \( g \), is certain that \( \alpha_a(g) = p \), that is, \( \beta_b(\alpha_a(g) = p|g) = 1 \), with \( p < 1 \). Also, Bob is certain after \( g \) that Ann expected him to accept the greedy offer with probability \( q \), that is, \( \beta_b(\alpha_a(g) = q|g) = 1 \). Finally, suppose Bob initially expected to get the fair offer \( (\alpha_b(f) = \)
1), so his frustration after \( g \) is \( F_b(a_1; \alpha_b) = 2 - 1 = 1 \). Then the extent of Bob’s blame on Ann’s intentions is

\[
B_{ba}(g; \beta_b) = \min \{2 - [2(1 - p) + qp], 1\} = \min \{p(2 - q), 1\}.
\]

If \( p \) is low enough, or \( q \) high enough, Bob does not blame all his frustration on Ann. He gives her some credit for the initial intention to make the fair offer with probability \( 1 - p > 0 \), and the degree of credit depends on \( q \).

5 Equilibrium analysis

We consider two notions of equilibrium. The first one is the sequential equilibrium (SE) concept of Battigalli & Dufwenberg (2009),\(^\text{13}\) extending Kreps & Wilson’s (1982) classic solution to psychological games. In a complete information framework like the one we adopt here for simplicity,\(^\text{14}\) SE requires that each player \( i \) is certain and never changes his mind about the true beliefs and plans, hence intentions, of his co-players. We find this feature quite unintuitive; therefore, we also explore a generalization – “polymorphic sequential equilibrium” (PSE) – that allows for meaningful updating about others’ intentions.

Battigalli & Dufwenberg’s (2009) SE concept gives equilibrium conditions for infinite hierarchies conditional probability systems. In our particular application, utility functions only depend on first- or second-order beliefs, so we define SEs for assessments comprising beliefs up to only the second order. Since, technically, first-order beliefs are features of second-order beliefs (see 2.2), we provide definitions that depend only on second-order beliefs, which gives SEs for games where psychological utility functions depend only of first-order beliefs as a special case.

5.1 Sequential equilibrium (SE)

Fix a game form and decision-utility functions \( u_i(h, \cdots) : A_i(h) \times \Delta_i^2 \rightarrow \mathbb{R} \) \((i \in I, h \in H)\). An assessment \( \) is a profile of behavioral strategies and

\(^{13}\)We consider the version for preferences with own-plan dependence and “local” psychological utility functions (see Battigalli & Dufwenberg 2009, Section 6).

\(^{14}\)Recall that complete information means that the rules of the game and players’ (psychological) preferences are common knowledge.
beliefs \((\sigma_i, \beta_i)_{i \in I} \in \times_{i \in I} \Sigma_i \times \Delta_i^2\) such that \(\Sigma_i = \times_{h \in H} \Delta(A_i(h))\) and for each \(i \in I\), \(\sigma_i\) is the plan \(\alpha_{i,i}\) entailed by CPS \(\beta_i:\)

\[
\sigma_i(a_i|h) = \alpha_{i,i}(a_i|h) = \beta_i \left( Z(h, a_i) \times \Delta_{-i}^\perp \right) \tag{8}
\]

for all \(i \in I\), \(h \in H\), \(a_i \in A_i(h)\). (8) implies that the behavioral strategies contained in an assessment are implicitly determined by players’ beliefs about paths; therefore, they could be dispensed with. We follow Battigalli & Dufwenberg (2009) and make behavioral strategies explicit in assessments only to facilitate understanding and comparisons with the equilibrium refinements literature.

**Definition 1** An assessment \((\sigma_i, \beta_i)_{i \in I}\) is consistent if, for all \(i \in I\), \(h \in H\), and \(a = (a_j)_{j \in I(h)} \in A(h)\),

(a) \(\alpha_i(a|h) = \times_{j \in I(h)} \sigma_j(a_j|h)\),
(b) \(\text{marg}_{\Delta_i} \beta_i(\cdot|h) = \delta_{\alpha_{-i}}\),

where \(\alpha_j\) is derived from \(\beta_j\) for each \(j \in I\), and \(\delta_{\alpha_{-i}}\) is the Dirac probability measure that assigns probability one to the singleton \(\{\alpha_{-i}\} \subseteq \Delta_{-i}^1\).

Condition (a) requires that players’ beliefs about actions satisfy independence across co-players (on top of own-action independence), and each \(i\) expects each \(j\) to behave as specified by \(j\)’s plan \(\sigma_j = \alpha_{j,j}\). Condition (b) requires that players’ beliefs about co-players’ first-order beliefs (hence their plans) are correct and never change, on or off the path.

**Definition 2** An assessment \((\sigma_i, \beta_i)_{i \in I}\) is a sequential equilibrium (SE) if it is consistent and satisfies the following sequential rationality condition: for all \(h \in H\) and \(i \in I(h)\)

\[
\text{Supp} \sigma_i(\cdot|h) \subseteq \arg \max_{a_i \in A_i(h)} u_i(h, a_i; \beta_i).
\]

Battigalli & Dufwenberg (2009) prove the following existence result: \(15\)

**Theorem 1** If \(u_i(h, \cdot; \cdot)\) is continuous for each \(i \in I\) and \(h \in H\) then there is at least one SE.

\(15\)See their Section 6 for comments about the extension of their existence theorem to games own-plan dependent preferences and local utility functions.
Since frustration and blame as defined above are continuous in beliefs, decision-utility is also continuous, and we obtain existence in all cases of interest:

**Corollary 1** Every game with SA, ABB, or ABI has at least one SE.

We close this section with three examples which combine to illustrate the nature of our SE concept at work (including a weakness) and that the notions of SA, ABB (both versions), and ABI may produce different predictions.

**Example 8** Consider Figure C. With $u^{ABB}_a$ (either version) or $u^{ABI}_a$ Andy won’t blame Bob so his SE-choice is $N$ but with $u^{SA}_a$ Andy may choose $T$. Recall that $F_a(B; \alpha_a) = 2(1 - \varepsilon) + \varepsilon \alpha_a(N|B) - 1$, so Andy’s utility from $N$ and $T$ is

\[
\begin{align*}
    u^{SA}_a(B, N; \alpha_i) &= 1 - \theta^{SA}_a[2(1 - \varepsilon) + \varepsilon \alpha_a(N|B) - 1] \cdot 1, \\
    u^{SA}_a(B, T; \alpha_i) &= 0 - \theta^{SA}_a[2(1 - \varepsilon) + \varepsilon \alpha_a(N|B) - 1] \cdot 0 = 0.
\end{align*}
\]

It follows from sequential rationality of SE that one possibility is $\alpha_a(N|B) = 1$ and $u^{SA}_a(B, N; \alpha_i) \geq u^{SA}_a(B, T; \alpha_i)$ implying $\theta^{SA}_a \leq \frac{1}{1 - \varepsilon}$. Another possibility is $\alpha_a(N|B) = 0$ and $u^{SA}_a(B, N; \alpha_i) \leq u^{SA}_a(B, T; \alpha_i)$ implying $\theta^{SA}_a \leq \frac{1 - \varepsilon}{2\varepsilon}$. If $\theta^{SA}_a \in \left(\frac{1}{1 - \varepsilon}, \frac{1 - 2\varepsilon}{1 - \varepsilon}\right)$, we can solve for an SE where $u^{SA}_a(B, N; \alpha_i) = u^{SA}_a(B, T; \alpha_i)$ and $\alpha_a(N|B) = \frac{1}{\varepsilon \theta^{SA}_a} - \frac{1 - 2\varepsilon}{\varepsilon}$; the more prone to get angry Andy is the more likely he believes he will take it out on Bob, so the less he expects initially and the less frustrated he is when $B$ happens. This illustrates how we cannot take for granted that an SE exists where players use degenerate plans (a point relevant also for $u^{ABB}_i$ or $u^{ABI}_i$ in other games).

**Example 9** Consider Figure A. Can $(U, L)$ be part of a SE? The answer is yes under ABI and the blaming-unexpected-deviations version of ABB. To see this note that Ann and Bob act as-if selfish (as they are not frustrated). Hence they’d deviate if they could gain material payoff. In the SE, they’d expect 5 if not deviating, making Ann the sole deviation candidate (she’d get 6 > 5 were Penny to choose $P$; for Bob, 5 is the best he could hope for). Ann deviating can be dismissed though, since if $(D, L)$ were reached Penny would not blame Bob (the only co-player she can punish) under either relevant blame function, and so she would choose $N$ (regardless of $\theta_p$). Under the could-have-been version of ABB, however, it may be impossible to sustain
a SE with \((U, L)\); at \((D, L)\) Penny would blame each of Ann and Bob (as explained earlier). By choosing \(P\) she hurts Bob more than she helps Ann and would do so if

\[
u_p^{ABB}((D, L), P; \alpha_p) > u_p^{ABB}((D, L), N; \alpha_p)
\]

\[
0 - 6\theta_p^{ABB}B_{pa}((D, L); \alpha_p) > 1 - 8\theta_p^{ABB}B_{pa}((D, L); \alpha_p)
\]

The rhs of the last inequality uses \(B_{pa}((D, L); \alpha_p) = B_{pa}((D, L); \alpha_p)\). Since \(B_{pa}((D, L); \alpha_p) = F_p((D, L); \alpha_1) = 1 > 0\), Penny would choose \(P\) if \(-6\theta_p^{ABB} > 1 - 8\theta_p^{ABB} \iff \theta_p^{ABB} > 1/2\), so Ann would want to deviate and choose \(D\).

**Example 10** Consider Figure B. Every utility function discussed admits \((g, y)\) as a SE, regardless of anger sensitivity; if Bob expects \(g\) he cannot be frustrated so when asked to play he maximizes material payoff. Under SA and ABB (both versions), \((f, n)\) qualifies as another SE if \(\theta_b \geq \frac{1}{3}\); following \(g\) Bob would be frustrated and choose \(n\), so Ann chooses \(f\). Under ABI \((f, n)\) cannot be an SE. To verify, assume it were, so \(\alpha_a(f) = 1\). Since the SE concept does not allow for players revising belief about beliefs we get \(\beta_b(\alpha_a(f) = 1|g) = 1\) and \(B_{ba}(g; \beta_b) = 0\); Bob maintains his belief that Ann planned to to choose \(f\), hence she intended to maximize Bob’s payoff. Hence, Bob would choose \(y\) (and, predicting this, Ann would choose \(g\)), contradicting that \((f, n)\) is a SE. Next, note that \((g, n)\) is not a SE under any concept; given SE beliefs Bob wouldn’t be frustrated and so he would choose \(y\). The only way to get rejected offers with positive probability in a SE is with non-degenerate plans. To find such a SE, note that we need \(\alpha_a(g) \in (0, 1)\); if \(\alpha_a(g) = 0\) Bob wouldn’t be reached and if \(\alpha_a(g) = 1\) he wouldn’t be frustrated (and, hence he would choose \(n\)). Since Ann uses a non-degenerate plan she must be indifferent, so \(\alpha_a(y) = 2/3\), implying that Bob is indifferent too. In SE, Bob’s frustration is \([2(1 - \alpha_a(g)) + \frac{2}{3}\alpha_a(g) - 1] = [1 - \frac{4}{3}\alpha_a(g)]^+\), which equals his blame of Ann under SA and ABB. Letting \(\theta_b\) be his anger sensitivity, we get the indifference condition

\[
1 - \theta_b[1 - \frac{4}{3}\alpha_a(g)]^+ \cdot 3 = 0 - \theta_b[1 - \frac{4}{3}\alpha_a(g)]^+ \cdot 0
\]

\[
\iff \alpha_a(g) = \frac{3}{4} - \frac{1}{4\theta_b} \text{ where } \theta_b > \frac{1}{3}
\]
The more prone to anger Bob is the more likely he is to get the low offer, so Bob’s initial expectations, and hence his frustration and blame, is kept low. Under ABI we get another indifference condition:

\[ 1 - \theta_b B_{ba}(g; \beta_b) \cdot 3 = 0 - \theta_b B_{ba}(g; \beta_b) \cdot 0 \]

\[ \iff \]

\[ 1 - \theta_b \min \left\{ 1 - \frac{4}{3} \alpha_a(g), \frac{4}{3} \alpha_a(g) \right\} \cdot 3 = 0 \]

The left term in brackets is Bob’s frustration while \( \frac{4}{3} \alpha_a(g) = 2 - [2(1-\alpha_a(g)) + \frac{2}{3} \alpha_a(g)] \) is the difference between the maximum payoff Ann could plan for Bob and her actually planned one. The first term is lower if \( \alpha_a(g) \geq \frac{3}{8} \), so if we can solve the equation for such a number we duplicate the SA/ABB-solution; again, this is doable if \( \theta_b > \frac{1}{3} \). If \( \theta_b \geq \frac{2}{3} \), with ABI, there is second non-degenerate equilibrium plan with \( \alpha_a(g) \in (0, \frac{2}{3}) \) such that \( \alpha_a(g) = \frac{1}{4\theta_b} \); to see this, solve the ABI indifference condition assuming that \( \frac{4}{3} \alpha_a(g) \leq \frac{1}{2} - \frac{4}{3} \alpha_a(g) \).

This SE exhibits starkly different comparative statics: the more prone to anger Bob is the less likely he is to get a low offer and the less he blames Ann following \( g \) in light of her intention to choose \( f \) with higher probability.

In the last example we explained why with ABI \((f, n)\) cannot be an SE. We find the interpretation unappealing. If Bob initially expects Ann to choose \( f \), and she doesn’t, so that Bob is frustrated, then he would rate her choice a mistake and not blame her! It may seem more plausible for Bob not to be so gullible, and instead revise his beliefs of Ann’s intentions. The SE concept rules that out. Because of this, and because it makes sense regardless, we next define an alternative concept that to a degree overcomes the issue.

### 5.2 Polymorphic sequential equilibrium (PSE)

Suppose a game is played by agents drawn at random and independently from large populations, one for each player role \( i \in I \). Different agents in the same population \( i \) have the same belief-dependent preferences, but they may have different plans, hence different beliefs about paths, even if their beliefs agree about the behavior and beliefs of co-players \(-i\). In this case, we say that the population is “polymorphic.” Once an agent playing in role \( i \)

\footnote{We are not modelling incomplete information.}
observes some moves of co-players, he makes inferences about the intentions of the agents playing in the co-players’ roles.

Let $\lambda_i$ be a finite support distribution over $\Sigma_i \times \Delta_i$, with $\text{Supp}\lambda_i = \{(\sigma_{t_i}, \beta_{t_i}), (\sigma_{t_i}', \beta_{t_i}') \ldots \}$. We interpret $\lambda_i$ as a statistical distribution of plans and beliefs of agents playing in role $i$ and, for every $(\sigma_{t_i}, \beta_{t_i}) \in \text{Supp}\lambda_i$, we let $\lambda_{t_i}$ denote the fraction of agents in population $i$ with plan and beliefs $(\sigma_{t_i}, \beta_{t_i})$.\textsuperscript{17} We refer to such index $t_i$ as a “type” of $i$.\textsuperscript{18} Also, we denote by

$$T_i(\lambda_i) = \{t_i : (\sigma_{t_i}, \beta_{t_i}) \in \text{Supp}\lambda_i\}$$

the set of possible types of $i$ in distribution $\lambda_i$. Thus, we can write $\lambda_i = ((\sigma_{t_i}, \beta_{t_i}))_{t_i \in T_i(\lambda_i)}$. Also, we write $T_{-i}(\lambda_{-i}) = \times_{j \neq i} T_j(\lambda_j)$ for the set of profiles of co-players’ types.

Let us take the perspective of an agent of type $t_i$ who knows that the distribution over co-players’ types is $\lambda_{-i} = \prod_{j \neq i} \lambda_j$ and believes that the behavior of each $t_j$ is indeed described $t_j$’s plan $\sigma_{t_j}$ (in principle, $t_i$ may otherwise believe that $t_j$ behaves differently from his plan). Then it is possible to derive the conditional probability of a type profile $t_{-i}$ given history $h$.

Given that beliefs satisfy independence across players (everybody knows that there is independent random matching), the distribution is independent of $t_i$ and can be factorized. In the current two-stage setting we have:

$$\lambda_{-i}(t_{-i}|\emptyset) = \prod_{j \neq i} \lambda_j(t_j|\emptyset) = \prod_{j \neq i} \lambda_{t_j},$$

and

$$\lambda_{-i}(t_{-i}|a^1) = \frac{\prod_{j \neq i} \sigma_{t_j}(a_{j}^1)\lambda_{t_j}}{\sum_{\nu_{-i} \in T_{-i}(\lambda_{-i})} \prod_{j \neq i} \sigma_{t_j}(a_{j}^1)\lambda_{t_j}}$$

$$= \frac{\prod_{j \neq i} \sigma_{t_j}(a_{j}^1)\lambda_{t_j}}{\prod_{j \neq i} \sum_{t_j' \in T_j(\lambda_j)} \sigma_{t_j'}(a_{j}^1)\lambda_{t_j'}}$$

$$= \prod_{j \neq i} \frac{\sigma_{t_j}(a_{j}^1)\lambda_{t_j}}{\sum_{t_j' \in T_j(\lambda_j)} \sigma_{t_j'}(a_{j}^1)\lambda_{t_j'}},$$

for all $t_{-i}$ and $a^1$, provided that $\sum_{t_j'} \sigma_{t_j'}(a_{j}^1)\lambda_{t_j'} > 0$ for each $j \neq i$. Letting

$$\lambda_j(t_j|a^1) = \frac{\sigma_{t_j}(a_{j}^1)\lambda_{t_j}}{\sum_{t_j' \in T_j(\lambda_j)} \sigma_{t_j'}(a_{j}^1)\lambda_{t_j'}},$$

\textsuperscript{17}The marginal of $\lambda_i$ on $\Sigma_i$ is a behavior strategy mixture (see Selten 1975).

\textsuperscript{18}They are “types” in the sense of epistemic game theory (e.g. Battigalli et al. 2013).
we get
\[ \lambda_{-i}(t_{-i}|a^1) = \prod_{j \neq i} \lambda_j(t_j|a^1). \]
We say that \( \lambda_j \) is **fully randomized** if \( \sigma_{t_j} \) is strictly positive for every type \( t_j \in T_j(\lambda_j) \). If each \( \lambda_j \) is fully randomized, then, for all \( h \in H \),
\[ \lambda_{-i}(t_{-i}|h) = \prod_{j \neq i} \lambda_j(t_j|h). \]

**Definition 3** A **polymorphic assessment** is a profile of finite support probability measures \( \lambda = (\lambda_i)_{i \in I} \in \times_{i \in I} \Delta(\Sigma_i \times \Delta_i^2) \) such that, for every \( i \in I \) and \( t_i \in T_i(\lambda_i) \), \( \sigma_{t_i} \) is the behavior strategy obtained from \( \beta_{t_i} \) as per (8). A polymorphic assessment \( \lambda \) is **consistent** if there is a sequence \( (\lambda^n)_{n=1}^{\infty} \) of polymorphic assessments converging to \( \lambda \) such that, for all \( j \in I \) and \( n \in \mathbb{N} \), \( \lambda^n_j \) is fully randomized, and
\((a-p)\) for all \( h \in H \), \( a \in A(h) \), and \( t_i \in T_i(\lambda^n_i) \),
\[ \alpha^n_{t_i,-i}(a_{-i}|h) = \prod_{j \neq i} \sum_{t_j \in T_j(\lambda^n_j)} \sigma^n_{t_j}(a_j|h)\lambda^n_j(t_j|h). \]
\((b-p)\) for all \( h \in H \) and \( t_i \in T_i(\lambda^n_i) \),
\[ \text{marg}_{\Delta_{-i}^l} \beta^n_{t_i}(\cdot|h) = \sum_{t_{-i} \in T_{-i}(\lambda^n_{-i})} \lambda^n_{-i}(t_{-i}|h)\delta_{\alpha^n_{t_{-i}}}, \]
where, for all \( j \in I \), \( t_j \in T_j(\lambda^n_j) \) and \( n \in \mathbb{N} \), \( \alpha^n_j \) is the first-order CPS derived from \( \beta^n_{t_j} \).

Condition (a-p) extends the independence (a) of Definition 1 to the multiple-types setting. Condition (b-p) implies that, conditional on the co-players’ types, everyone has correct beliefs about the beliefs of others, including their plans, but uncertainty about co-players’ types allows for uncertainty and meaningful updating about such beliefs. Conditions (a-p) and (b-p) imply that different types of the same player share the same beliefs about co-players, but may have different plans. Therefore, Definition 3 is a minimal departure from the notion of consistent assessment, allowing for uncertainty and meaningful updating about the plans, hence intentions, of co-players.
Definition 4 A polymorphic assessment $\lambda$ is a **polymorphic sequential equilibrium (PSE)** if it is consistent and satisfies the following sequential rationality condition: for all $h \in H$, $i \in I(h)$, and $t_i \in T_i(\lambda_i)$,

$$\text{Supp}_i(\cdot|\cdot) \subseteq \arg\max_{a_i \in A_i(h)} u_i(h, a_i; \beta_i).$$

Remark 2 Every SE is a degenerate (or monomorphic) PSE. Therefore, Theorem 1 implies that, if every decision-utility function $u_i(h, \cdot; \cdot)$ ($i \in I$, $h \in H$) is continuous, then there is at least one PSE. In particular, every game with SA, ABB, or ABI has at least one PSE.

We finally demonstrate how the PSE alters predictions in the ultimatum mini-game.

Example 11 Consider Figure B. Everything we said about SE remains relevant with PSE in the sense that if $|\text{Supp}\lambda_i| = 1$ for all $i$ then the SE analysis is a special case. Interesting new possibilities arise if $|\text{Supp}\lambda_a| = 2$ though. First, let us note something which holds for the SA and ABB (both versions) utility functions: Recall that in an SE with non-degenerate plans we had $\alpha_a(g) = \frac{3}{4} - \frac{1}{4\theta_b}$ (with $\theta_b > \frac{1}{3}$) to keep Bob indifferent. Suppose instead there are two types of Ann, a fraction of $\frac{3}{4} - \frac{1}{4\theta_b}$ of them planning to choose $g$ while the others plan for $f$. There is a corresponding PSE where (naming Ann’s types by planned choice) $\lambda_a = (u, v, z, w)$, $\alpha_f(y|g) = \alpha_g(y|g) = \alpha_a(y|g) = \frac{2}{3}$, and this holds for also for ABI, not only SA and ABB. The first-order belief of type $f$ of Ann is derived from $\beta_f$, etc. Bob initially believes Ann is either an $f$-type or a $g$-type, assigning probability $\frac{3}{4} - \frac{1}{4\theta_b}$ to the latter possibility. Following $g$ he ceases to assign positive probability to being matched with an $f$-type, assigning instead probability 1 to the $g$-type, a meaningful form of updating about Ann’s intentions implied by consistency (Def. 3). This inference makes ABI work exactly as ABB (and SA). Bob’s frustration is as in Example 10, so equal to his blame of Ann for each blaming function. So again Bob is indifferent between $y$ and $n$, and sequentially rational if $\alpha_b(y|g) = \frac{2}{3}$. Condition $\alpha_f(y|g) = \alpha_g(y|g) = \frac{2}{3}$ implies that both types of Ann are indifferent, hence sequentially rational. Thus, starting with the non-degenerate SE under ABB (and SA) we obtained a PSE, under every blaming function, where Ann is purified. Second, consider the previous SE of Example 10 under ABI where
\(\alpha_a(g) = \frac{1}{3}\theta_b \geq \frac{2}{3}\). Now there is no corresponding PSE with Ann-purification. To see this, note that whereas before Bob’s blame of Ann were \(\lambda_a = ((f, \beta_f), (g, \beta_g))\) the corresponding expression would now be \(\lambda_a = ((f, \beta_f), (g, \beta_g))\) putting us back in the world where \(\alpha_b(g) = \alpha_a(g) = \frac{3}{4} - \frac{1}{4}\theta_b\). Third, it is impossible to purify Bob in this example; we are stuck with \(j\) supp \(b\) = 1. To see this, suppose instead that \(b = ((y, \beta_y), (n, \beta_n))\), \(\alpha_y(y|g) = 1\), and \(\alpha_n(n|g) = 0\). This cannot be a PSE as neither of Bob’s types would exhibit sequential rationality. Type \(y\), because of his plan to accept, initially has higher expectations than type \(n\), who plans to reject, so type \(y\) is more frustrated and more blaming than type \(n\). If one type is sequentially rational, the other would want to match his choice, violating the own plan.

The second observation of the previous example can be generalized to all leader-followers games:

**Remark 3** Consider a leader-followers game. Every SE under ABB/SA where the behavioral strategy of the leader has full support corresponds to a PSE under ABI and ABB/SA (with the same parameter values) where the leader is purified.

**Proof** Let \((\sigma_i, \beta_i)_{i \in I}\) be an SE under ABB/SA with parameter profile \((\theta_i)_{i \in I}\), and suppose that Supp\(\sigma_i(\varnothing)\) = \(A_i(\varnothing)(\varnothing)\). Construct a polymorphic consistent assessment \(\tilde{\lambda}\) as follows: for each follower \(i\), \(T_i(\lambda_i) = \{t_i\}\) (a singleton) and \((\tilde{\sigma}_i, \tilde{\beta}_i) = (\sigma_i, \beta_i)\). For the leader \(\iota(\varnothing)\), \(T_i(\lambda_i(\sigma)) = A_i(\sigma)(\varnothing)\), and, for each type \(a_i(\varnothing)\), \(\tilde{\sigma}_{a_i(\varnothing)}(a_i(\varnothing)\varnothing) = 1\) and \(\tilde{\alpha}_{a_i(\varnothing)}(a_i(\varnothing)\varnothing) = \sigma(a_i)\) for all non-terminal \(a^1\), where \(\tilde{\alpha}_{a_i(\varnothing)}\) is the first-order belief derived from \(\tilde{\beta}_{a_i(\varnothing)}\). By construction, each type of leader is indifferent, because the leader (who acts as-if selfish) is indifferent in assessment \((\sigma_i, \beta_i)_{i \in I}\). As for the followers, they have the same first-order beliefs, hence the same second-stage frustrations as in \((\sigma_i, \beta_i)_{i \in I}\). Under ABB/SA, blame always equals frustration in leader-followers games. As for ABI, Bayes’ rule implies that, after observing \(a^1 = a_i(\varnothing)\), each follower becomes certain that the leader indeed planned to choose \(a_i(\varnothing)\) with probability one, and blame equals frustrations in this case too. Therefore, the incentive conditions of the followers hold in \(\tilde{\lambda}\) as in \((\sigma_i, \beta_i)_{i \in I}\) for all kinds of decision utility (ABI, ABB, SA) under the same parameter profile \((\theta_i)_{i \in I}\).

\(^{19}\)Recall that, by Remark 1, SA is equivalent to both versions of ABB in such games.
6 Multistage extension

To be written

7 Discussion

Incorporating the effects of emotions in economic analysis is a balancing act. One wants to focus on sentiments that make empirical sense, but human psychology is multi-faceted and there is no unambiguous yardstick. Our chosen formulation provides a starting point for exploring how anger shapes economic interaction, and experimental or other evidence will help to assess empirical relevance and suggest revised formulas. We conclude by discussing topics that one way or another may be helpful for gaining perspective on, building on, or further developing our work. It is a mixture of commentary on chosen concepts, comparisons with related notions in the literature, and remarks about empirical tests.

Frustration  Our theory is restrictive in how we model sources and effects of frustration. Regarding sources, recall our discussion of Cases 4-5, and the idea that frustration may be grounded in negative surprises other than those concerning own material reward (e.g. fairness, regret).

As regards the effects of frustration, we considered changes to a player’s utility function but we neglected other plausible adjustments. Gneezy & Imas (2014) report data from an intriguing experiment involving two-player zero-sum material payoff games. In one game players gain if they are powerful, in the other they are rewarded for being smart. Gneezy & Imas explore an added game-feature; before play starts, one subject may frustrate his opponent and force him to hang around in the lab to do boring tasks after the play ends. A thus frustrated player’s performance is enhanced when strength is beneficial (possibly from increased adrenalin flow), but reduced in the game where cool logic is called for (as if an angered player becomes cognitively impaired). Our model captures neither consideration.

Valence and action-tendency  Psychologists classify emotions in multiple ways. Two prominent aspects are valence, or the value the decision-maker associates with a sentiment, and action-tendency, or how behavior is shaped as the sentiment occurs. Both notions may, in principle, have
bearing on anger. For example, anger may have negative valence, say if a frustration-laden life is taxing or decreases longevity. Perhaps such considerations steer people to avoid frustrations, say by not investing in the stock market. That said, the distinguishing feature of anger that psychologists tend to stress most concerns its action-tendency of aggression, not its valence. In developing our theory, we have exaggerated this, abstracting away from valence altogether while emphasizing aggression. This is reflected in the decision utility functions, which are shaped by current frustration, but there is no notion of valence. Anticipated future frustrations do not affect decision utility.\textsuperscript{20}

\textbf{Blame} We explored various ways a player may blame the frustration he experiences on others, yet more notions are conceivable. For example, with anger from blaming behavior player $i$'s blame of $j$ depends on what $i$ believes he would truly get at counterfactual histories, not on the most he could get there were he not to vent any anger there. We would defend this modeling choice as a reflection of local agency and abstracting away from valence; $i$'s current agent views other agents as uncontrollable, and he has no direct care for their frustrations.

Another example relates to how we model anger from blaming intentions. $i$'s blame of $j$ depends on $\beta_i$, his second-order beliefs. Recall that the interpretation concerns beliefs about beliefs about material payoffs. It does not concern beliefs about beliefs about frustration, which would be third-rather than second-order beliefs. Battigalli & Dufwenberg (2007), in a context which concerned guilt rather than anger, worked with such a third-order belief based notion of blame.

Our blame notions one way or another assess marginal impact of other player. For example, consider a game where $i$ exits a building while each $j \in I \setminus \{i\}$, unexpectedly to $i$, simultaneously hurls a bucket of water at $i$, who thus gets soaked. According to our blame notions, $i$ cannot blame $j$ as long as there are at least two hurlers. One could imagine alternatives where $i$ blames, say, all the hurlers on the grounds that collectively they could have thwarted $i$’s misery.

\textsuperscript{20}In previous work we modeled another emotion, namely guilt – see, e.g., Battigalli & Dufwenberg (2007), Chang et al. (2011). To gain further perspective one may note that in that work our approach to valence and action-tendency was the opposite. Guilt may have valence (negative!) as well as action-tendency (say to engage in “repair behavior”; see e.g. Silfver (2007). In modeling guilt we highlighted valence while neglecting action-tendency.
Several recent experiments have explored interesting aspects of blame (Bartling & Fischbacher 2012, Gurdal et al. 2014, Celen et al. 2014). Is a field of the economics-of-blame emerging? We wish to emphasize that our focus on blame is restricted to its relation to frustration only. Our paper is not about blame more generally, as of course there are many reasons besides frustration that may lead people to blame each other.\footnote{For example, Celen et al. (2014) present a model where \(i\) asks himself how he would have behaved had he been in \(j\)'s position and had \(j\) beliefs, and \(i\) blames \(j\) if \(j\) appears to be less generous to \(i\) than \(i\) would have been. This involves that a player may blame another even if he is not surprised/frustrated. Relatedly, one may imagine a model where players blame those they consider unkind, as defined in reciprocity theory (cf. the subsubsection below), again something independent of frustration.}

Kőszegi & Rabin Card & Dahl show that reports of domestic abuse go up when football home teams favored to win lose. They informally argue that this is in line with Kőszegi & Rabin’s (2006, 2007) theory of expectations-dependent reference points. One way to think of our paper is as providing a formalization of that idea. Kőszegi & Rabin model the loss felt when a player gets less than he expected, which one may think of as a form disappointment with negative valence (cf. Bell 1985, Loomes & Sugden 1986). That account per se does not imply that aggression follows, but it may seem natural to add such an angle. In other words, one may imagine a general model where disappointment/frustration has negative valence as well as an action-tendency of aggression. Kőszegi & Rabin’s model would be the modification that looks at the negative valence part only. Our model of simple anger focuses on the action tendency only, which is enough to capture the effect the Card & Dahl found.\footnote{Modeling details distinguish how we measure frustration from how Kőszegi & Rabin measure loss, in ways that are not central to our discussion here (e.g. concerning how we cap frustration at the highest attainable as opposed to actual payoff).}

Anger management People aware of their inclination to be angry may attempt to ‘manage’ or ‘contain’ it. Our players anticipate how frustrations shape the behavior of themselves and others, and they may avoid or seek certain subgames because of that. However, there are other interesting related phenomena that we do not address: can player \(i\) somehow adjust \(\theta^x_i\) (where \(x \in \{SA, ABB, ABI\}\)) say by taking an “anger management class”? If so, would rational individuals want to raise, or to lower, their \(\theta^x_i\)? How might
that depend on the game forms they play? These are potentially relevant questions related to how we have modeled action-tendency in this paper. Further issues would arise if we were to add some valence aspects of anger.

**Rotemberg’s approach** In a series of intriguing papers Rotemberg explores how consumer anger shapes firms’ pricing (2005, 2011), as well as interaction in ultimatum games (2008). He proposes (versions of) a theory in which players are slightly altruistic, and consumers/responders also care about their co-players’ degrees of altruism. Namely, they abruptly become very angry and punish a co-player whom they come to believe has an altruism parameter lower than some (already low) threshold. “One can thus think of individual $i$ as acting as a classical statistician who has a null hypothesis that people’s altruism parameter is at least as large as $\lambda$. If a person acts so that $i$ is able to reject this hypothesis, individual $i$ gains ill-will towards this person” (Rotemberg 2008, p. 424).

On the one hand, as a literal statement of what makes people upset, this assumption does not match well our reading of the psychology of anger. Recall that these are anchored in goal-blockage, where individual are unexpectedly denied things they care about. Matters like “own payoff,” “fairness,” “quality of past decisions” come to mind; a co-player’s altruism being $\lambda$ rather than $\lambda - \varepsilon$, where both $\lambda$ and $\varepsilon$ are tiny numbers, hardly does. On the other hand, it is impressive how well Rotemberg’s model captures the action in his data sets. It is natural to wonder whether our models could achieve that too. As regards behavior in ultimatum (and some other) games, there is already some existing evidence that is seemingly in line with our modeling efforts; see the discussion in the final subsection below. As regards understanding pricing, we leave for empirical economists that task of exploring how our models might fare if applied to Rotemberg’s data sets.

**Negative reciprocity** Negative reciprocity (cf. Rabin 1993, Dufwenberg & Kirchsteiger 2004, Falk & Fischbacher 2006, Sebald 2010) joins anger as a form of motivation that can trigger hostile action. In some cases implications may be similar. However, anger and negative reciprocity differ in key ways and it is instructive to point out how. The following sketched comparison is with Dufwenberg & Kirchsteiger’s notion of sequential reciprocity equilibrium (SRE; refer to their article for formal definitions):

First, in the hammering-one’s-thumb game form of Figure C, Andy may
take it out on Bob if he is motivated by simple anger. If he were motivated by reciprocity this could never happen: Andy’s kindness, since he is a dummy-player, equals 0, implying that a reciprocal Bob chooses as-if selfish. In this example reciprocity captures intuitions similar to the ABI and ABB concepts, as perceived kindness assesses intentions similarly to how blame is apportioned.

Second, that analogy only carries so far, however. A player may be perceived as unkind even if he fails to hurt another, whereas under all our anger notions frustration is a prerequisite for hostile response. The following game illustrates:

![Figure D. Failed attack](image)

If \( b \) is asked to play then \( a \)’s attack failed. Under reciprocity theory (suitably augmented to allow incorporating a chance move; cf. Sebald 2010), \( b \) would deem \( a \) unkind, and if sufficiently motivated by reciprocity choose \( p \) in response. By contrast, under either of our anger concepts (SA, ABB, ABI) \( b \) would not be frustrated, and since frustration is a prerequisite for hostility \( b \) would choose \( n \).

Third, reciprocity allows for so-called “miserable equilibria,” where a player reciprocates expected unkindness before it occurs. For example, in the mini-ultimatum game of Figure B, \((g, n)\) may be a SRE. Ann makes the greedy offer \( g \) despite believing that Bob will reject, because given her beliefs about Bob’s beliefs, she perceives Bob as seeing this coming, which
makes him unkind, so she punishes him by choosing $g$. This pattern of self-fulfilling prophecies of destructive behavior has no counterpart under either of our anger notions. Since Ann moves at the root, she cannot be frustrated, and hence, regardless of how prone to anger she may be in terms of anger sensitivity, she chooses as-if selfish.

Fourth, with reference to our discussion of cooling-off effects in section 6, these have no counterpart in Dufwenberg & Kirchsteiger’s theory, which rather makes the same prediction in the games of Figures B and D. Reciprocal players do not cool off, they say things like “la vengeance est un plat qui se mange froid.”

**Experimental testing**  Our models tell various stories of how interaction may play out when players prone to anger interact. It is natural to wonder about empirical relevance, and here experiments may be helpful. We would like to offer several related remarks.

First, several existing experimental studies provide evidence in favor of the notion that emotions drive behavior, and that many of them, and anger in particular, are generated from comparisons of outcomes with expectations: A number of studies find evidence for anger as the driving force behind costly punishment. A few papers rely on emotion-self reports: Pillutla & Murnigham (1996) find that reported anger predicted rejections better than perceived unfairness in ultimatum games. Fehr & Gächter (2002) elicited self-reports of the level of anger towards free riders in a public goods game, concluding that negative emotions including anger are the proximate cause of costly punishment in the game. A few studies directly connect unfulfilled expectations and costly punishment in ultimatum games. Schotter & Sopher (2007) measure second-mover expectations in ultimatum games, concluding that unfulfilled expectations drive rejections of low offers. Similarly, Sanfey (2009) finds that psychology students who are told that a typical offer in the ultimatum game is $4-$5 reject low offers more frequently than students who are told that a typical offer is $1-$2. A series of papers by Frans van Winden and coauthors records both emotions and expectations in the power-to-take game (which resemble ultimatum games, but allow for partial rejections). These papers show that second-mover expectations about first-mover take rates in power-to-take games are a key factor in the decision to destroy income. Furthermore, they find that anger-like emotions are triggered by the

\[^{23}\text{Bosman & van Winden (2002), Bosman et al. (2005), Reuben & van Winden (2008).}\]
difference between expected and actual take rates. In these experiments the
difference between the take rate and the reported fair take rate is not sig-
nificant in determining anger-like emotions, suggesting that deviations from
expectations, rather than from fairness benchmarks, drive both anger and
the destruction of endowments in the games. Reuben & van Winden also
argue that existing models of belief-dependent reciprocity miss an important
trigger of anger by focusing on equilibrium predictions where actions are cor-
rectly anticipated. In our models anger arises only off the equilibrium path,
addressing Reuben & van Winden’s point that anger is triggered by unful-
filled expectations. Finally a literature in neuroscience connects expectations
and social norms to study the neural underpinnings of emotional behavior.
In Xiang et al. (2013), subjects respond to a sequence of ultimatum game
offers whilst undergoing fMRI imaging. Unbeknownst to subjects, the ex-
perimenter controls the distribution of offers in order to manipulate beliefs.
The authors find that rejections occur more often when subjects expect high
offers relative to when they expect low offers. They make an important
connection between norm violations and reward prediction errors from rein-
forcement learning, which are known to be the computations instantiated by
the dopaminergic reward system. Xiang et al. note that “when the expecta-
tion (norm) is violated, these error signals serve as control signals to guide
choices. They may also serve as the progenitor of subjective feelings.”

Second, going forwards it would seem useful to develop tests specifically
designed to target key features of our theory. For example, which version
– SA, ABB, ABI – seems more empirically relevant, and how does the an-
swer depends on context details (e.g., is SA perhaps more relevant for tired
subjects?). Some insights may again be gleaned from existing studies. For
example, Gurdal et al. (2014) study games where an agent invests on behalf
of a principal, choosing between a safe outside option and a risky alternative.
If the latter is chosen then it turns out that many principals punish the agent
if and only if by chance a poor outcome is realized. This may seem to indi-
cate some empirical relevance of our ABB solution (relative to ABI). That
said, Gurdal et al.’s intriguing design is not tailored to specifically test our
theory (and beliefs and frustrations are thus not measured), so more work
seems needed to draw clearer conclusions.

Finally, one must keep in mind that our models are abstractions. We theo-
rize about the consequences of anger while neglecting myriad other obviously
important aspects of human motivation (say altruism, warm glow, inequity
aversion, reciprocity, social status, or other emotions like guilt, disappoint-
ment, regret, or anxiety). Our models are not intended to explain every data pattern but rather to highlight the would-be consequences of anger, if anger were the only form of motivation at play (on top of some concern for own material payoff, of course). This statement may seem trivially obvious, but it has subtle implications for how to evaluate experimental work. To illustrate, consider again the Failed Attack game form in Figure D and suppose that in an experiment many subjects in player b’s position chose to punish (p). It would be silly to say that this constitutes a rejection of our theory (which predicts n rather than p), as what may obviously be going on is that one of the important forms of motivation that our theory deliberately abstracts away from is affecting subjects choices (presumably negative reciprocity, in line with our observations in the previous subsection). It would be sensible to ask, however, if those choices of p were in fact driven by anger (as might be measured say by emotions self-reports, or hormonal activity) and if they were (as opposed to being driven only by alternative motivations that we abstract away from, like negative reciprocity) then that would indicate that our theory could benefit from revision.

References


