# Frustration and Anger in Games*

Pierpaolo Battigalli† Martin Dufwenberg‡ Alec Smith§¶

October 19, 2018

**Abstract**

Frustration, anger, and blame have important consequences for economic and social behavior, concerning for example monopoly pricing, contracting, bargaining, 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 strategic settings.

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

*JEL* codes: C72, D01, D91

*This paper modifies and extends Smith (2009). Pierpaolo Battigalli gratefully acknowledges financial support from ERC advanced grant 324219. We thank Chiara Aina, Federico Bobbio, Roberto Corrao, Giacomo Lanzani, Paolo Leonetti, Paola Moscariello, and Marco Stenborg Petterson for excellent research assistance, and Doug Bernheim, Steve Brams, Leda Cosmides, Vince Crawford, Nicodemo De Vito, Uri Gneezy, Pierfrancesco Guarino, Michele Griessmair, Heike Hennig-Schmidt, Botond Kőszegi, Alex Imas, Joshua Miller, David Rietzke, Julio Rotemberg, Emanuel Vespa, several reviewers, and many seminar and conference audiences for helpful comments.

†Bocconi University and IGIER. Email: pierpaolo.battigalli@unibocconi.it.

‡University of Arizona, University of Gothenburg, CESifo. Email: martind@eller.arizona.edu.

§Virginia Tech, Department of Economics. Email: alecsmith@vt.edu.

¶Corresponding author.

# 1 Introduction

Anger can shape economic outcomes. Consider three cases:

> **Case 1:** In 2015 Turing Pharmaceuticals raised the price of Daraprim, a therapeutic drug, from $12 to $750 per dose. The company was subsequently accused of price gouging. Should Turing have considered the consequences of customer anger before determining the new price for the drug?
>
> **Case 2:** When local football teams favored to win instead lose, the police get more reports of husbands assaulting wives (Card & Dahl 2011). Do unexpected losses spur vented frustration?
>
> **Case 3:** Following the sovereign debt crises that began in 2009, some EU countries embarked on austerity programs. Was it because citizens lost benefits that some cities experienced riots?

Pricing, domestic violence, political landscapes: these are important themes, and we propose that others (involving—say—recessions, contracting, arbitration, terrorism, road rage, or support for populist political candidates) could plausibly be imagined. However, to carefully assess the impact of anger on social and economic interactions, one needs a theory that predicts outcomes based on the decision-making of anger-prone individuals and that accounts for strategic considerations. We develop such a theory.

Insights from psychology about the triggers and repercussions of anger are evocative. The behavioral consequences of emotions are called "action tendencies," and the action tendency associated with anger is aggression and the urge to retaliate. Angry players may be willing to forgo material gains to punish others, or be predisposed to aggression when this serves as a credible threat. 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 and social implications. We take the basic insights about anger that psychology has produced as input and inspiration for our theory.[1]

We study the strategic interaction of decision makers who become angry when they are frustrated.[2] Frustration occurs when someone is unexpectedly denied something he or she cares about. We assume that people are frustrated when they get less material rewards than

[1] The psychology literature is huge. A source of inspiration is *International Handbook of Anger* (Potegal *et al.* 2010) offering a cross-disciplinary perspective reflecting "affective neuroscience, business administration, epidemiology, health science, linguistics, political science, psychology, psychophysiology, and sociology" (p. 3). The absence of "economics" in the list may indicate that our approach is original!

[2] A large body of work in psychology connects frustration, anger, and aggression, beginning with Dollard *et al.* (1939). See, for example, Averill (1982), Berkowitz (1989), and the (*op. cit.*) *Handbook*, especially the chapters by Lewis, Wranik & Scherer, and Berkowitz.

expected.[3] They then become hostile towards whomever they blame. Because a player's frustration depends on his beliefs about others' choices, and the blame a player attributes to another may depend on his beliefs about others' choices or beliefs, all our models find their intellectual home in the framework of psychological game theory; see Geanakoplos, Pearce & Stacchetti (1989) and Battigalli & Dufwenberg (2009) (B&D).

In our model, initially expected material payoffs are the reference point to which outcomes are compared to generate frustration. This modeling choice mirrors that of several other behavioral models, including Kőszegi & Rabin's (2006, 2007) model of reference-dependent preferences, guilt aversion (Dufwenberg 2002, Battigalli & Dufwenberg 2007), and earlier models of disappointment aversion (Bell 1985, Loomes & Sugden 1986). While this approach may not incorporate every aspect of frustration discussed in the psychology literature, it is consistent with a broad range of behavioral phenomena, allowing for frustration and anger to be belief-dependent and capturing the stylized fact that costly punishment involves violations of consequentialism (*e.g.* Falk *et al.* 2003, 2008).

There are a number of ways to model the assignment of blame.[4] We present three approaches that result in distinct utility functions. Players motivated by simple anger (SA) become generally hostile when frustrated. In contrast, those motivated by anger from blaming behavior (ABB) or anger from blaming intentions (ABI) are more discriminating, asking who caused, or who intended to cause, their frustration. SA captures the well-known psychological phenomenon of **displaced aggression**, where an angry person takes out frustration on a blameless bystander.[5] However, for some authors, blame or other-responsibility is a prerequisite for anger.[6] ABB and ABI are consistent with this view.

We develop and apply a modeling framework in which players have beliefs about both others' beliefs and actions as well as their own actions. In any game form first-movers are never frustrated, and therefore behave (at the root) as if to maximize their material payoffs given their beliefs. We define and establish the existence of a notion of sequential equilibrium (SE) that adapts to our framework the one developed in B&D. In pure-strategy sequential equilibria, frustration arises only off the equilibrium path, and furthermore, in generic perfect information game forms, there is always an equilibrium with anger that is realization-equivalent to the material-payoff equilibrium, though anger may also result in additional equilibria. In two-player *leader-follower* games, followers with SA, ABB, or ABI always do at least as well,

[3]That frustration depends on expectations is well supported in the psychology literature, e.g. Berkowitz (1978, p. 697): "Unlike deprivations, frustrations can only be surprising (to a greater or lesser extent). People who do not expect to reach their goals are not anticipating the pleasure these goals would bring. Their hopes are not dashed if they have no hopes." Our focus on material rewards is admittedly restrictive. See the Discussion in Section 7.

[4]See e.g. Alicke (2000), Battigalli & Dufwenberg (2007), and Halpern (2016, Chapter 6).

[5]See Marcus-Newhall *et al.* (2000).

[6]See e.g. the chapter by Wranik & Scherer in the (*op. cit.*) *Handbook*.

materially, as players who are not anger-prone. Taking into account that evolutionary pressure is driven by material payoffs (e.g., Buss 2016), this result is consistent with the work of Sell, Cosmides, & Tooby (2009), who argue that anger is the result of a process of natural selection for behaviors that resolve bargaining conflicts in favor of the anger-prone individual. We also formally develop the notion of a **threat** in order to provide a partial characterization of SE with anger: the presence of threats allows anger-prone followers to obtain more than in the material-payoff equilibrium and give less to the leader, while their absence implies that equilibria with SA, ABB, or ABI are equivalent to the material payoff equilibrium.

We show via examples how our models can encapsulate Cases 1 and 2 above (for Case 3, cf. Passarelli & Tabellini 2017). Case 1 is captured by an ultimatum minigame, where SA and ABB allow for a pure SE involving rejection of the greedy offer, and all our concepts allow for an SE where rejection occurs with positive probability. Case 2 is modeled with SA in an example that we call "Hammering one's thumb."[7] In contrast, incorporating notions of blame into our analysis of this situation via either ABB and ABI eliminates displaced aggression.

Finally, we illustrate how the effects of anger are sensitive to assumptions about how rapidly the reference expectation that determines frustration is updated. We show that our model can be interpreted either via the notion of "fast play," where the reference belief is fixed at the initial history, or "slow play," where the reference expectation changes each period. Slow play formalizes the phenomenon of cooling-off, whereby anger subsides over time (Grimm & Mengel, 2011; Oechssler *et al.*, 2015).

A small literature examines the role of anger in economic behavior. Selten (1978) discusses the implications of the frustration-aggression hypothesis of Dollard *et al.* (1939) for behavior in the chain store game, though he does not develop a formal model.[8] Other earlier work exploring the role of anger in solving commitment problems includes Hirshleifer (1987), Frank (1988), and Elster (1998). Most recent studies are empirical or experimental, indicative of hostile action occurring in economic situations, based on either observational data or experimental data.[9] A few studies present theories different from ours, including Rotemberg (2005, 2008, 2011), Brams (2011), Winter (2014), Winter *et al.* (2016), Akerlof (2016), and Passarelli & Tabellini (2017). We compare and contrast our approach with these and with models of

[7]The example is inspired by Frijda (1993), who says "Many experiences or responses of anger... are elicited by events that involve no blameworthy action" and suggests that simple frustrations such as "one's car refusing to start, finding one's bicycle has a flat tyre, rain on the fifth day of one's holiday after four previous days of rain... hitting one's head on the kitchen shelf, dropping a needle for the third time in a row" or "hammering one's thumb" may result in anger. He goes on to say that "the target of anger may be a person who has fallen ill on the day of one's party, or one who just happened to be present when a plan failed."

[8]The chain store stage game is strategically equivalent to the ultimatum minigame.

[9]See Anderson & Simester (2010) and Rotemberg (2005, 2011) on pricing; Card & Dahl (2011) and Munyo & Rossi (2013) on violence; Carpenter & Matthews (2012), Gurdal *et al.* (2014), Gneezy & Imas (2014), Persson (2018), van Leeuwen *et al.* (2018), Aina *et al.* (2018), and Dufwenberg *et al.* (2018*a*,*b*) for experiments.

distributional preferences and reciprocity in Section 6.

Our approach differs from the previous literature in that we do not start with data, but with notions from psychology which we incorporate into general games, and we are led to use assumptions which differ substantially from previous theoretical work. We develop most of our analysis for a two-stage setting described in Section 2. Section 3 defines frustration, blame, anger, and utility. Section 4 examines equilibria. Section 5 generalizes our approach to multistage games and develops notions of fast and slow play. Section 6 compares our approach to other related models. Section 7 concludes. Proofs are collected in the Appendix.

# 2 Preliminaries

In this section we develop a framework well-suited to the study of frustration, anger and blame. We first describe the rules of interaction (the game form), then first- and second-order conditional belief systems that concern a player's system of beliefs regarding the behavior and beliefs of others, as well as about own actions which we refer to as a player's plan.

## 2.1 Game form

Consider a finite two-stage game form describing the rules of interaction and the consequences of players' actions. The set of players is $I$. To ease notation, we assume that all players take actions simultaneously at each stage. Thus, nodes are histories $h$ of action profiles $a^t = (a_i^t)_{i \in I}$; $h = \varnothing$ is the empty history (the root), $h = (a^1)$ a history of length one, which may be terminal or not, and $h = (a^1, a^2)$ a history of length 2, which is terminal. $H$ is the set of nonterminal histories and $Z$ is the set of terminal histories (end nodes). The set of feasible actions of $i$ given $h \in H$ is $A_i(h)$. This set is a singleton if $i$ is not active given $h$. Thus, for all $h \in H$, $I(h) = \{i \in I : |A_i(h)| > 1\}$ is the set of active players given $h$. In a **perfect information game**, $I(h)$ is a singleton for each $h \in H$. We omit parentheses whenever no confusion may arise. For example, we may write $h = a^1$ instead of $h = (a^1)$, and $h = (a_i^1, a_j^2)$ if $i$ (resp. $j$) is the only first (resp. second) mover. Finally, we let $A(h) = \times_{i \in I} A_i(h)$ and $A_{-i}(h) = \times_{j \neq i} A_j(h)$. The material consequences of players' actions are determined by a profile of monetary payoff functions $(\pi_i : Z \to \mathbb{R})_{i \in I}$. We say that a *perfect information* game has **no relevant ties** if distinct terminal histories yield different payoffs for the player who is active at the longest common prefix.[10] For two-stage games, this means that different actions of the first mover lead to different material payoffs for the first mover, and different actions of a second mover

[10] See Battigalli (1997). It can be checked that the no-relevant-ties (NRT) property is *generic* with respect to material payoff functions $\pi \in \mathbb{R}^{Z \times I}$: the closure of the set of $\pi$ that do not satisfy NRT has Lebesgue measure 0 in $\mathbb{R}^{Z \times I}$.

lead to different material payoffs for this second mover.[11] 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, let $I_c = I \cup \{c\}$, and the sets of first and second movers may include $c$: $I(\varnothing), I(a^1) \subseteq I_c$. If the chance player is active at $h \in H$, its move is described by probability mass function $\sigma_c(\cdot|h) \in \Delta(A_c(h))$.

The following example, to which we will return in our discussion of blame, is here employed to illustrate our notation:

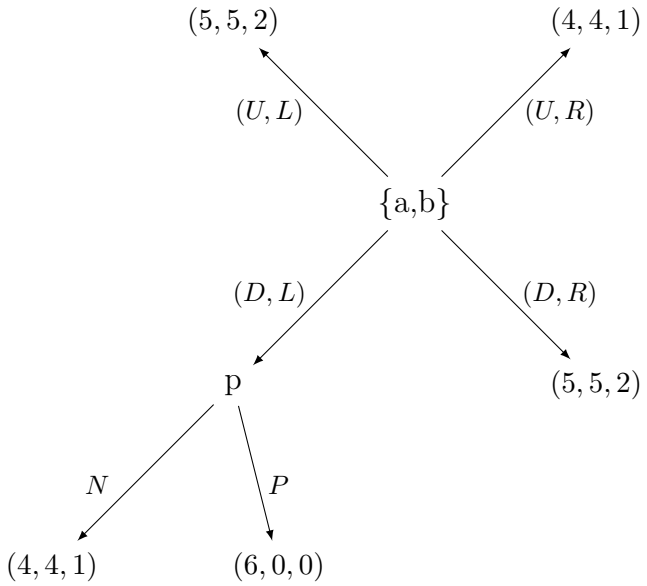


**Figure A.** Asymmetric punishment.

**Example 1** Ann and Bob (a and b in Fig. A) move simultaneously in the first stage. Penny the punisher (p) may move in the second stage; by choosing $P$ she increases $\pi_a$ and decreases $\pi_b$. Profiles of actions and monetary payoffs are listed in players' alphabetical order. We have:

$$H = \{\varnothing, (D, L)\}\,,\ Z = \{(U, L), (U, R), (D, R), ((D, L), N)\,, ((D, L), P)\}\,,$$
$$I(\varnothing) = \{a, b\},\ I\left((D, L)\right) = \{p\},$$
$$A_a(\varnothing) = \{U, D\},\ A_b(\varnothing) = \{L, R\},\ A_p\left((D, L)\right) = \{N, P\}.\quad \blacktriangle$$

[11] Who is the second mover may depend on the action of the first mover.

## 2.2 Beliefs

It is conceptually useful to distinguish three aspects of a player's beliefs: beliefs about co-players' actions, beliefs about co-players' beliefs, and the player's plan, which we represent as beliefs about own actions. Beliefs are defined conditional on each history. Abstractly denote by $\Delta_{-i}$ the space of co-players' beliefs (the formal definition is given below). Player $i$'s beliefs can be compactly described as conditional probability measures over paths and beliefs of others, *i.e.*, over $Z \times \Delta_{-i}$. Events, from $i$'s point of view, are subsets of $Z \times \Delta_{-i}$. Events about behavior take form $Y \times \Delta_{-i}$, with $Y \subseteq Z$; events about beliefs take form $Z \times E_{\Delta_{-i}}$, with $E_{\Delta_{-i}} \subseteq \Delta_{-i}$.[12]

**Personal histories** To model how $i$ determines the subjective value of feasible actions, we add to the commonly observed histories $h \in H$ also personal histories of the form $(h, a_i)$, with $a_i \in A_i(h)$. 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 can determine the value of $a_i$ using his beliefs conditional on this event ($i$ knows in advance how he is going to update his beliefs conditional on what he observes). We denote by $H_i$ the set of histories of $i$—standard and personal—and by $Z(h_i)$ the set of terminal successors of $h_i$.[13] 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 $h$. Note that $h \prec h'$ implies $Z(h') \subseteq Z(h)$, with strict inclusion if at least one player (possibly $c$) is active at $h$.

**First-order belief systems** For each $h_i \in H_i$, player $i$ holds beliefs $\alpha_i(\cdot|Z(h_i)) \in \Delta(Z(h_i))$ about the actions that will be taken in the continuation of the game. The system of beliefs $\alpha_i = (\alpha_i(\cdot|Z(h_i)))_{h_i \in H_i}$ must satisfy two properties. First, the rules of conditional probabilities hold whenever possible: if $h_i \prec h_i'$ then for every $Y \subseteq Z(h_i')$

$$\alpha_i(Z(h_i')|Z(h_i)) > 0 \Rightarrow \alpha_i(Y|Z(h_i')) = \frac{\alpha_i(Y|Z(h_i))}{\alpha_i(Z(h_i')|Z(h_i))}. \tag{1}$$

[12] $\Delta_{-i}$ turns out to be a compact metric space. Events are Borel measurable subsets of $Z \times \Delta_{-i}$. We do not specify terminal beliefs of $i$ about others' beliefs, as they are not relevant for the models in this paper.

[13] That is, $H_i = H \cup \{(h, a_i) : h \in H, i \in I(h), a_i \in A_i(h)\}$. The definition of $Z(h_i)$ is standard for $h_i \in H$; for $h_i = (h, a_i)$ we have $Z(h, a_i) = \bigcup_{a_{-i} \in A_{-i}(h)} Z(h, (a_i, a_{-i}))$.

We use obvious abbreviations to denote conditioning events and the conditional probabilities of actions: for all $h \in H$, $a = (a_i, a_{-i}) \in A_i(h) \times A_{-i}(h)$,

$$\begin{aligned}
\alpha_i(a|h) &= \alpha_i\left(Z(h,a)|Z(h)\right), \\
\alpha_{i,i}(a_i|h) &= \sum_{a'_{-i} \in A_{-i}(h)} \alpha_i(a_i, a'_{-i}|h), \\
\alpha_{i,-i}(a_{-i}|h) &= \sum_{a'_i \in A_i(h)} \alpha_i(a'_i, a_{-i}|h).
\end{aligned}$$

Note that $\alpha_{i,i}(a_i|h) = \alpha_i\left(Z(h,a_i)|Z(h)\right)$, and that (1) implies $\alpha_i\left(a^1, a^2|\varnothing\right) = \alpha_i\left(a^2|a^1\right)\alpha_i(a^1|\varnothing)$.

With this, we can write in a simple way our second requirement, that $i$'s beliefs about the actions simultaneously taken by the co-players are independent of $i$'s action: for all $h \in H$, $i \in I$, $a_i \in A_i(h)$, and $a_{-i} \in A_{-i}(h)$,

$$\alpha_{i,-i}(a_{-i}|h) = \alpha_{i,-i}(a_{-i}|h, a_i). \tag{2}$$

Properties (1)–(2) imply

$$\alpha_i(a_i, a_{-i}|h) = \alpha_{i,i}(a_i|h)\alpha_{i,-i}(a_{-i}|h).$$

Thus, $\alpha_i$ is made of two parts, what $i$ believes about own behavior *and* about the behavior of others. The array of probability measures $\alpha_{i,i} \in \times_{h \in H}\Delta\left(A_i(h)\right)$ is—technically speaking—a behavior 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 $\alpha_{i,-i} \in \times_{h \in H}\Delta\left(A_{-i}(h)\right)$ corresponds to a behavior strategy. With multiple co-players, $\alpha_{i,-i}$ corresponds instead to a "correlated behavior strategy." Whatever the case, $\alpha_{i,-i}$ gives $i$'s conditional beliefs about others' behavior, and these beliefs may not coincide with the plans of others. We emphasize: a player's plan does not describe actual choices, actions on the path of play are the only actual choices.

A system of conditional probability measures $\alpha_i = \left(\alpha_i(\cdot|Z(h_i))\right)_{h_i \in H_i}$ satisfying (1)–(2) is a **first-order** belief of $i$; let $\Delta_i^1$ denote the space of such. It can be checked that $\Delta_i^1$ is a compact metric space, hence the same holds for $\Delta_{-i}^1 = \times_{j \neq i}\Delta_j^1$, the space of co-players' first-order beliefs profiles.

**Second-order belief systems** Players do not only hold beliefs about paths, they also hold beliefs about the beliefs of co-players. In the following analysis, the only co-players' beliefs affecting the values of actions are their first-order beliefs. Therefore, we limit our attention to **second-order** beliefs, *i.e.*, systems of conditional probability measures $\left(\beta_i(\cdot|h_i)\right)_{h_i \in H_i} \in$

$\times_{h_i \in H_i} \Delta\left(Z(h_i) \times \Delta^1_{-i}\right)$ that satisfy properties analogous to (1)–(2).[14] First, if $h_i \prec h'_i$ then

$$\beta_i(h'_i|h_i) > 0 \Rightarrow \beta_i\left(E|h'_i\right) = \frac{\beta_i\left(E|h_i\right)}{\beta_i(h'_i|h_i)} \tag{3}$$

for all $h_i, h'_i \in H_i$ and every event $E \subseteq Z(h'_i) \times \Delta^1_{-i}$. Second, $i$'s choice cannot influence co-players' first-order beliefs and simultaneous choices, so $i$'s beliefs satisfy an independence property:

$$\beta_i\left(Z\left(h,(a_i,a_{-i})\right) \times E_\Delta|(h,a_i)\right) = \beta_i\left(Z\left(h,(a'_i,a_{-i})\right) \times E_\Delta|(h,a'_i)\right), \tag{4}$$

for every $h \in H$, $a_i, a'_i \in A_i(h)$, $a_{-i} \in A_{-i}(h)$, and event $E_\Delta \subseteq \Delta^1_{-i}$ about co-players' first-order beliefs. The space of $i$'s second-order beliefs is denoted by $\Delta^2_{-i}$.

It can be checked that given $\beta_i \in \Delta^2_i$ and $\alpha_i\left(Y|h_i\right) = \beta_i\left(Y \times \Delta^1_{-i}|h_i\right)$ for all $h_i \in H_i$ and $Y \subseteq Z$, we obtain a system $\alpha_i$ satisfying (1)–(2), *i.e.*, an element of $\Delta^1_i$. This $\alpha_i$ is the first-order belief implicit in $\beta_i$. Whenever we write in a formula beliefs of different orders for $i$, we assume that $\alpha_i$ is derived from $\beta_i$, otherwise beliefs of different orders would not be mutually consistent. Also, we write initial beliefs omitting the empty history, as in $\beta_i\left(E\right) = \beta_i\left(E|\varnothing\right)$ or $\alpha_i(a) = \alpha_i(a|\varnothing)$, whenever this causes no confusion.

**Conditional expectations** Let $\psi_i$ be any real-valued measurable function of variables that $i$ does not know, *e.g.*, the terminal history or the co-players' first-order beliefs. Then $i$ can compute the expected value of $\psi_i$ conditional on any history $h_i \in H_i$ by means of his belief system $\beta_i$, denoted $\mathbb{E}[\psi_i|h_i;\beta_i]$. If $\psi_i$ depends only on actions, *i.e.*, on the path $z$, then $\mathbb{E}[\psi_i|h_i;\beta_i]$ is determined by the $\alpha_i$ derived from $\beta_i$, and we can write $\mathbb{E}[\psi_i|h_i;\alpha_i]$. In particular, $\alpha_i$ gives the conditional expected material payoffs:

$$\begin{aligned}
\mathbb{E}[\pi_i|h;\alpha_i] &= \sum_{z \in Z(h)} \alpha_i(z|h)\pi_i(z), \\
\mathbb{E}[\pi_i|\left(h,a_i\right);\alpha_i] &= \sum_{z \in Z(h,a_i)} \alpha_i(z|h,a_i)\pi_i(z)
\end{aligned}$$

for all $h \in H$, $a_i \in A_i(h)$. $\mathbb{E}[\pi_i|h;\alpha_i]$ is what $i$ expects to get conditional on $h$ given $\alpha_i$, which also specifies $i$'s plan. $\mathbb{E}[\pi_i|(h,a_i);\alpha_i]$ is $i$'s expected payoff of action $a_i$. If $a_i$ is what $i$ planned to choose at $h$, $\alpha_{i,i}(a_i|h) = 1$, and then $\mathbb{E}[\pi_i|h;\alpha_i] = \mathbb{E}[\pi_i|\left(h,a_i\right);\alpha_i]$. For initial beliefs, we omit $h = \varnothing$ from such expressions; in particular, $i$'s initially expected material payoff is $\mathbb{E}[\pi_i;\alpha_i]$.

Table 1 summarizes our modeling setup.

[14] We use obvious abbreviations, such as writing $h$ for event $Z(h) \times \Delta^1_{-i}$, whenever this causes no confusion.

| Notation | Terminology |
|---|---|
| $i \in I$ | players |
| $h \in H$ | non-terminal, or partial histories |
| $I(h) \subseteq I$ | set of active players at $h$ |
| $t \in \{1, 2\}$ | stages, or periods |
| $A_i(h)$, $A(h)$, $A_{-i}(h)$ | set of actions and action profiles at $h$ |
| $a_i^t$ | action of $i$ in stage $t$ |
| $a^t$ $(a_{-i}^t)$ | action profile (of others) in stage $t$ |
| $\sigma_i(a_i^t \vert h)$ | behavior strategies |
| $z \in Z$ | terminal histories |
| $Z(h)$ | terminal successors of $h$ |
| $\pi_i : Z \to \mathbb{R}$ | monetary payoff function of $i \in I$ |
| $\alpha_i, \alpha_{-i}, \alpha$ | First-order beliefs and belief profiles |
| $\beta_i, \beta_{-i}, \beta$ | Second-order beliefs and belief profiles |

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

# 3 The frustration-aggression hypothesis, anger, and blame

## 3.1 Frustration

Anger is triggered by frustration. While we focus on anger as a social phenomenon—frustrated players blame, become angry with, and care for the payoffs of *others*—our account of frustration refers to own payoffs only. In Section 7 (in hindsight of definitions to come) we discuss this approach in depth.

We define player $i$'s **frustration** at history $h$, given his first-order belief system $\alpha_i$ as

$$\mathrm{F}_i(h; \alpha_i) = \left[ \mathbb{E}[\pi_i; \alpha_i] - \max_{a_i \in A_i(h)} \mathbb{E}[\pi_i | (h, a_i); \alpha_i] \right]^+,$$

where $[x]^+ = \max\{x, 0\}$. In words, frustration is given by the gap, if positive, between $i$'s initially expected payoff and the currently best expected payoff he believes he can obtain. Diminished expectation—$\mathbb{E}[\pi_i | h; \alpha_i] < \mathbb{E}[\pi_i; \alpha_i]$—is only a necessary condition for frustration. For $i$ to be frustrated it must also be the case that $i$ cannot close the gap.

At the root, frustration must always be zero, because nothing has yet happened and so expectations cannot be diminished:

**Remark 1** *For every player $i \in I$ and system of first-order beliefs $\alpha_i \in \Delta_i^1$, frustration must equal 0 at the initial history $h = \varnothing$, since (1) and (2) imply*

$$\mathbb{E}\left[\pi_i; \alpha_i\right] = \sum_{a_i^1 \in A_i(\varnothing)} \alpha_{i,i}\left(a_i^1 | \varnothing\right) \mathbb{E}\left[\pi_i | a_i^1; \alpha_i\right] \leq \max_{a_i^1 \in A_i(\varnothing)} \mathbb{E}\left[\pi_i | a_i^1; \alpha_i\right].$$

Frustration is possible at the end nodes, but can't influence subsequent choices as the game is over. One might allow the anticipation of frustration to be felt at end nodes to influence earlier decisions; however, the assumptions we make below rule this out. In our 2-stage setting, all behaviorally relevant frustration occurs in the second stage and is given by

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

## 3.2 Simple Anger

Preferences over actions at a given node depend on expected material payoffs and frustration. A frustrated player is motivated 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 cognitive appraisals of blame. In general, player $i$ moving at history $h$ chooses action $a_i$ to maximize the expected value of a belief-dependent "decision utility" of the form

$$u_i\left(h, a_i; \beta_i\right) = \mathbb{E}\left[\pi_i | \left(h, a_i\right); \alpha_i\right] - \theta_i \sum_{j \neq i} \mathrm{B}_{ij}\left(h; \beta_i\right) \mathbb{E}\left[\pi_j | \left(h, a_i\right); \alpha_i\right], \tag{5}$$

where $\alpha_i$ is the first-order belief system derived from second-order belief $\beta_i$, and $\theta_i \geq 0$ is a sensitivity parameter. Thus, $\mathrm{B}_{ij}\left(h; \beta_i\right) \geq 0$ measures how much of $i$'s frustration is blamed on $j$, and the presence of $\mathbb{E}\left[\pi_j | \left(h, a_i\right); \alpha_i\right]$ in the formula translates this into a tendency to hurt $j$.

We assume that

$$\mathrm{B}_{ij}(h; \beta_i) \leq \mathrm{F}_i(h; \alpha_i). \tag{6}$$

Therefore, first-movers behave at the root (given their beliefs) as if they are motivated only by material self-interest, because there is no frustration in the first stage:

**Remark 2** *The decision utility of a first-mover coincides with expected material payoff:*

$$u_i\left(\varnothing, a_i; \beta_i\right) = \mathbb{E}[\pi_i | a_i; \alpha_i].$$

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

$$u_i\left(h, a_i; \beta_i\right) = \pi_i\left(h, a_i\right) - \theta_i \sum_{j \neq i} \mathrm{B}_{ij}\left(h; \beta_i\right) \pi_j\left(h, a_i\right).$$

We assume throughout that players' utilities are commonly known among them. This is partly a modeling choice made for reasons of analytical tractability, but we note that the approach is supported by recent intriguing experimental evidence reported by van Leeuwen *et al.* (2018). They find that (pre-play) "facial cues provide a credible signal of destructive behavior," and that subjects are to a degree capable of recognizing such "angry buttons."

Our most rudimentary hypothesis, **simple anger (SA)**, is that $i$'s tendency to hurt others is proportional to $i$'s frustration. SA is unmodulated by the cognitive appraisal (i.e., personal interpretation) of blame, so $\mathrm{B}_{ij}\left(h; \beta_i\right) = \mathrm{F}_i(h; \alpha_i)$:

$$u_i^{SA}(h, a_i; \alpha_i) = \mathbb{E}\left[\pi_i \middle| \left(h, a_i\right); \alpha_i\right] - \theta_i \sum_{j \neq i} \mathrm{F}_i(h; \alpha_i) \mathbb{E}\left[\pi_j \middle| \left(h, a_i\right); \alpha_i\right]. \tag{7}$$

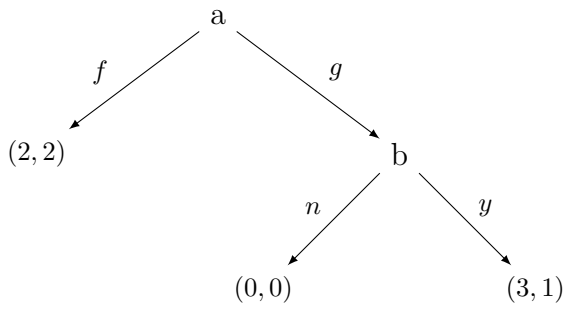


**Figure B.** Ultimatum Minigame.

We first demonstrate our model via the ultimatum minigame in Fig. B (Gale *et al.*, 1995). The game is a simplified version of the ultimatum game of Guth *et al.* (1982), and it has the same strategic structure as the stage game of Selten's (1978) chain-store game. As noted by Gale *et al.* (p. 76), models such as ours which imply that players will reject unfair offers in the ultimatum minigame "provide a possible resolution of the chain-store paradox that applies even in the case when there is just one potential entrant." Selten also considers that frustration and aggression may be relevant. Finally, vis-á-vis Case 1 in Section 1, the setup can also be interpreted as representing a monopoly seller (Ann) who can offer Bob either a high or a low split of the gains from trade.

**Example 2** *(Ultimatum Minigame)* Ann and Bob (a and b in Fig. B) negotiate: Ann can make a fair offer $f$, which is automatically accepted, or a greedy offer $g$, which Bob accepts or rejects. His frustration following $g$ is

$$\mathrm{F}_b(g;\alpha_b) = \left[(1-\alpha_b(g))\cdot 2 + \alpha_b(g)\alpha_b(y|g)\cdot 1 - 1\right]^+.$$

Therefore

$$u_b^{SA}(g,n;\alpha_b) - u_b^{SA}(g,y;\alpha_b) = 3\theta_b\left[2\left(1-\alpha_b(g)\right) + \alpha_b(g)\alpha_b(y|g) - 1\right]^+ - 1.$$

For Bob to be frustrated he must not expect $g$ with certainty. If frustrated, the less he expects $g$, and—interestingly—the less he plans to reject, the more prone he is to reject once $g$ materializes. The more resigned Bob is to getting a low payoff, the less frustrated and prone to aggression he is. Furthermore, it is readily seen from the example how our model can generate non-consequential behavior. Holding beliefs and other payoffs constant, increasing Bob's payoff from $f$ will lead to greater frustration after $g$, and so increases the disutility Bob receives from Ann's material payoff. This makes rejection (punishment of Ann) more attractive to Bob. ▲

Note that, in the example, when Bob rejects it is because he is truly angry and prefers $n$ to $y$. He is not trying signal his type to Ann in order to deter her from choosing $g$ in the future, marking a difference with "reputational models." Indeed, in the example there is no future behavior for Bob to influence.

Under SA a frustrated player goes after others rather indiscriminately. The word "rather" is justified because, as regards targets of aggression, our modeling of SA restricts attention to co-players, implicitly saying that persons who are not represented in a game are not targets. So the modeler has a responsibility to represent the appropriate environment. If another individual is a potential target (*e.g.* Bob's wife in addition to Ann), that person should be included (*e.g.* as a dummy player). The exact determination regarding whom to include in the description of the strategic environment is an empirical question. We have the qualitative idea that SA allows for innocent targets, and this is what we model. Future research may generate more nuanced insights.

We now move to consider models where targets of aggression must be less innocent than under SA.

## 3.3 Anger from blaming behavior (ABB)

Action tendencies may depend on a player's cognitive appraisal of how to blame others. When a frustrated player $i$ blames co-players for their behavior, he examines the actions chosen in

stage 1, without considering others' intentions: How much $i$ blames $j$ is defined by a continuous blame function that specifies blame $\mathrm{B}_{ij}(a^1;\alpha_i)$ that depends on $\alpha_i$ (but not $\beta_i$) such that

$$\mathrm{B}_{ij}(a^1;\alpha_i) = \begin{cases} 0, & \text{if } j \notin I(\varnothing), \\ \mathrm{F}_i(a^1;\alpha_i), & \text{if } \{j\} = I(\varnothing). \end{cases} \tag{8}$$

Eq. (8) is a complete specification of the $ij$-blame function if there is only one first mover; with two or more first movers, the second part is irrelevant, and the first just give a necessary condition. In particular, if $j$ is not active in the first stage, he cannot be blamed by $i$. If $j$ is the *only* active player, he is fully blamed.[15] We consider below specific forms of $\mathrm{B}_{ij}(h;\alpha_i)$ that satisfy (6) and (8). With this, $i$'s decision utility with **anger from blaming behavior (ABB)** is

$$u_i^{ABB}(h,a_i;\alpha_i) = \mathbb{E}\left[\pi_i|\left(h,a_i\right);\alpha_i\right] - \theta_i \sum_{j\neq i} \mathrm{B}_{ij}(h;\alpha_i)\mathbb{E}\left[\pi_j|\left(h,a_i\right);\alpha_i\right].$$

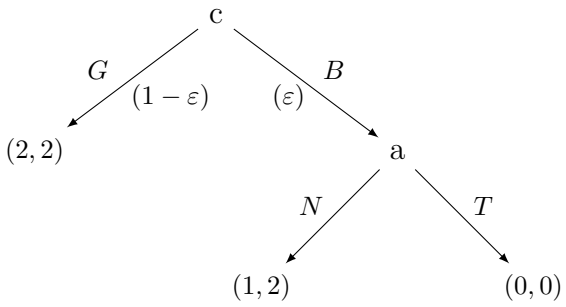


**Figure C.** Hammering one's thumb.

**Example 3** *(Inspired by Frijda 1993)* To illustrate the difference between SA and ABB, consider Fig. C. Andy the handyman (a) uses a hammer. His apprentice, Bob (b), is inactive. On a bad day (determined by chance, c) Andy hammers his thumb and can then take it out on Bob, or not. Assuming $\alpha_a(B) = \varepsilon < 1/2$, we have

$$\mathrm{F}_a(B;\alpha_a) = (1-\varepsilon)\cdot 2 + \varepsilon\alpha_a(N|B)\cdot 1 - 1 > 0.$$

With SA and with $\theta_a$ sufficiently high, on a bad day Andy chooses T in a fit of displaced aggression. But, since Bob is passive, with ABB Andy chooses $N$ regardless of $\theta_a$. ▲

Next, we contrast two specific functional forms for ABB.

[15] Recall that $I(h)$ is the set of active players at $h$, possibly including chance. For example, $I(\varnothing) = \{c\}$ in the game form of Fig. C.

**Could-have-been blame** When frustrated $i$ considers, for each $j$, what he would have obtained at most, in expectation, had $j$ chosen differently:

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

If this could-have-been payoff is more than what $i$ currently expects (that is, $\mathbb{E}[\pi_i | a^1; \alpha_i]$), then $i$ blames $j$, up to $i$'s frustration (so (6) holds):

$$\mathrm{B}_{ij}(a^1; \alpha_i) = \min\left\{\left[\max_{a'_j \in A_j(\varnothing)} \mathbb{E}\left[\pi_i | (a^1_{-j}, a'_j); \alpha_i\right] - \mathbb{E}[\pi_i | a^1; \alpha_i]\right]^+, \mathrm{F}_i(a^1; \alpha_i)\right\}. \tag{9}$$

(9) satisfies (8), in particular implying that $j$ cannot be blamed if he is not active in the first stage (cf. the multi-stage extension of could-have-been blame in Proposition 6 below).

**Example 4** Consider Penny at $a^1 = (D, L)$ in Fig. A. Her could-have-been payoff—wrt both Ann and Bob—is $2 \geq \mathbb{E}[\pi_p; \alpha_p]$, her updated expected payoff is $\mathbb{E}[\pi_p | (D, L); \alpha_p] \leq 1$, and her frustration is $[\mathbb{E}[\pi_p; \alpha_p] - 1]^+$. Therefore

$$\mathrm{B}_{pa}((D, L); \alpha_p) = \mathrm{B}_{pb}((D, L); \alpha_p) =$$

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

*i.e.*, each of Ann and Bob is fully blamed by Penny for her frustration. ▲

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

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

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

$$\mathrm{B}_{ij}(a^1; \alpha_i) =$$

$$\min\left\{\left[\sum_{a'_j \in A_j(\varnothing)} \alpha_{ij}(a'_j) \mathbb{E}\left[\pi_i | (a^1_{-j}, a'_j); \alpha_i\right] - \mathbb{E}[\pi_i | a^1; \alpha_i]\right]^+, \mathrm{F}_i(a^1; \alpha_i)\right\}. \tag{10}$$

If $j$ is not active in the first stage, we get

$$\mathrm{B}_{ij}(a^1; \alpha_i) = \min\left\{\left[\mathbb{E}[\pi_i | a^1; \alpha_i] - \mathbb{E}[\pi_i | a^1; \alpha_i]\right]^+, \mathrm{F}_i(a^1; \alpha_i)\right\} = 0;$$

since $j$ cannot have deviated, he cannot be blamed. If, instead, *only* $j$ is active in the first stage, then

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

and (10) yields full blame on $j$:

$$\mathrm{B}_{ij}(a^1; \alpha_i) = \min\left\{\left[\mathbb{E}[\pi_i; \alpha_i] - \mathbb{E}[\pi_i|a^1; \alpha_i]\right]^+, \mathrm{F}_i(a^1; \alpha_i)\right\} = \mathrm{F}_i(a^1; \alpha_i).$$

Therefore, like blame function (9), also (10) satisfies (8).

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

$$\mathrm{B}_{ij}(a^1; \alpha_i) = \min\left\{\left[\mathbb{E}[\pi_i|a^1; \alpha_i] - \mathbb{E}[\pi_i|a^1; \alpha_i]\right]^+, \mathrm{F}_i(a^1; \alpha_i)\right\} = 0;$$

$j$ did not deviate from what $i$ expected and is not blamed by $i$, marking a contrast to "could-have-been" blame (9).

**Example 5** Suppose that, in Fig. A, Penny is initially certain of $(U, L)$: $\alpha_p(U, L) = 1$ and $\mathbb{E}[\pi_p; \alpha_p] = 2$. Upon observing $(D, L)$ her frustration is $\mathrm{F}_p((D, L); \alpha_p) = [\mathbb{E}[\pi_p; \alpha_p] - 1]^+ = 1$. Using eq. (10), at $a^1 = (D, L)$, Penny fully blames Ann, who deviated from $U$ to $D$. Since

$$\sum_{a'_a \in A_a(\varnothing)} \alpha_{pa}(a'_a)\mathbb{E}\left[\pi_p|(a^1_{-a}, a'_a); \alpha_p\right] = \pi_p(U, L) = 2,$$

Penny's blame of Ann equals Penny's frustration

$$\mathrm{B}_{pa}((D, L); \alpha_p) = \min\left\{\left[2 - \mathbb{E}[\pi_p|a^1; \alpha_p]\right]^+, 1\right\} = 1.$$

Penny does not blame Bob, who played $L$ as expected. To see this, note that after $(D, L)$ Penny assesses how much she would have obtained had Bob behaved as expected:

$$\sum_{a'_b \in A_b(\varnothing)} \alpha_{pb}(a'_b)\mathbb{E}\left[\pi_p|(a^1_{-b}, a'_b); \alpha_p\right] = \mathbb{E}[\pi_p|(D, L); \alpha_p]$$

and

$$\mathrm{B}_{pb}((D, L); \alpha_p) = \min\left\{[\mathbb{E}[\pi_p|(D, L); \alpha_p] - \mathbb{E}[\pi_p|(D, L); \alpha_p]]^+, 1\right\} = 0,$$

in contrast to could-have-been blame under which Penny fully blames Bob (Example 4). ▲

In two-player games with a single leader and a single follower SA and ABB (both forms) are behaviorally equivalent (see Remark 4). However, in games with more than two followers, with chance moves, or with an inactive player in the second stage, SA and ABB give different predictions about behavior. We return to this issue and derive formal results in Section 4.3.

Next, we consider a more nuanced notion of blame, where players are concerned with their co-players' intentions, and preferences therefore depend upon second-order beliefs.

### 3.4 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 (5).

The maximum payoff that $j$, initially, can expect to give to $i$ is

$$\max_{a_j^1 \in A_j(\varnothing)} \sum_{a_{-j}^1 \in A_{-j}(\varnothing)} \alpha_{j,-j}(a_{-j}^1) \mathbb{E}\left[\pi_i | \left(a_j^1, a_{-j}^1\right); \alpha_j\right].$$

Note that

$$\begin{aligned} &\max_{a_j^1 \in A_j(\varnothing)} \sum_{a_{-j}^1 \in A_{-j}(\varnothing)} \alpha_{j,-j}(a_{-j}^1) \mathbb{E}\left[\pi_i | \left(a_j^1, a_{-j}^1\right); \alpha_j\right] \\ &\geq \sum_{a^1 \in A(\varnothing)} \alpha_j(a^1) \mathbb{E}\left[\pi_i | a^1; \alpha_j\right] = \mathbb{E}\left[\pi_i | \alpha_j\right], \end{aligned}$$

where the inequality holds by definition and the equality is implied by the chain rule (3). Note also that $\alpha_j(\cdot|a^1)$ is kept fixed under the maximization; we focus on what $j$ initially believes he could achieve, taking the view that at the root he cannot control $a_j^2$ but predicts his choice in stage 2. We assume that $i$'s blame of $j$ at $a^1$ equals $i$'s expectation, given $\beta_i$ and conditional on $a^1$, of the difference between the maximum payoff that $j$ can expect to give to $i$ and what $j$ actually plans/expects to give to $i$, capped by $i$'s frustration:

$$\mathrm{B}_{ij}(a^1; \beta_i) = \tag{11}$$

$$\min\left\{ \mathbb{E}\left[ \max_{a_j^1} \sum_{a_{-j}^1} \alpha_{j,-j}(a_{-j}^1) \mathbb{E}\left[\pi_i | \left(a_j^1, a_{-j}^1\right); \alpha_j\right] - \mathbb{E}[\pi_i; \alpha_j] \middle| a^1; \beta_i \right], \mathrm{F}_i(a^1; \alpha_i) \right\},$$

where $\alpha_i$ is derived from $\beta_i$. The expression is nonnegative as per the previously highlighted inequality. Now, $i$'s decision utility after $h = a^1$ is

$$u_i^{ABI}(h, a_i; \beta_i) = \mathbb{E}\left[\pi_i | (h, a_i); \alpha_i\right] - \theta_i \sum_{j \neq i} \mathrm{B}_{ij}(h; \beta_i) \mathbb{E}\left[\pi_j | (h, a_i); \alpha_i\right],$$

with $\mathrm{B}_{ij}(a^1; \beta_i)$ given by (11).

**Example 6** Return to Fig. B. The maximum Ann can expect to give Bob is 2, independently of $\alpha_a$. Suppose Bob, upon observing $g$, is certain Ann planned to offer $g$ with probability

$p < 1$: $\beta_b(\alpha_a(g) = p|g) = 1$. Also, Bob is certain after $g$ that Ann expected him to accept with probability $q$: $\beta_b(\alpha_a(y|g) = q|g) = 1$. Finally, suppose Bob initially expected to get the *fair* offer ($\alpha_b(f) = 1$), so his frustration after $g$ is $F_b(a^1; \alpha_b) = 2 - 1 = 1$. We get

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

If $p$ is low enough, or $q$ high enough, Bob does not blame all frustration on Ann. She gets some credit for initial intention to choose $f$ with probability $1 - p > 0$, and the credit depends on $q$. ▲

# 4 Equilibrium analysis

While in this paper we depart from traditional game-theoretic analysis in using belief-dependent decision utility, our analysis is otherwise traditional. We adapt B&D's sequential equilibrium (SE) concept.[16] For simplicity, we consider a complete information framework where the rules of the game and players' (psychological) preferences are common knowledge.[17] We interpret an SE as a profile of strategies and beliefs representing a "commonly understood" way to play the game by rational (utility maximizing) agents. Our approach allows us to investigate the implications of our belief-dependent utility model within a standard framework. This is a choice of focus more than an endorsement of SE as a solution concept.[18]

## 4.1 SE definition

The SE concept gives equilibrium conditions for infinite hierarchies of conditional probability systems. In our particular application, utility functions only depend on first- or second-order beliefs, so we define SE for assessments comprising beliefs up to only the second order. Since, technically, first-order beliefs are features of second-order beliefs (see Section 2), we provide definitions that depend only on second-order beliefs, which give SEs for games where

[16] B&D extend Kreps & Wilson's (1982) classic notion of SE to psychological games; here we consider the version (B&D, Section 6) for preferences with own-plan dependence and "local" psychological utility functions. A more subtle difference with B&D is that they assume to be common knowledge that players behave as planned, whereas we separate plans from behavior, and let the consistency of behavior with plan be a rationality condition.

[17] For an equilibrium analysis of incomplete-information psychological games see Attanasi, Battigalli & Manzoni (2016); for a non-equilibrium analysis see Battigalli, Charness & Dufwenberg (2013).

[18] 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 questionable. B&D (Sections 2, 5) argue that, with belief-dependent preferences, alternatives to SE like rationalizability, forward induction, and self-confirming equilibrium are even more plausible than with standard preferences.

psychological utility functions depend only on first-order beliefs as a special case. Finally, although so far we have restricted our analysis of frustration and anger to two-stage game forms, our abstract definitions of equilibrium for games with belief-dependent preferences (and the associated existence theorem) apply to all multistage game forms.

Fix a game form and decision utility functions $u_i(h,\cdot;\cdot) : A_i(h) \times \Delta_i^2 \to \mathbb{R}$ $(i \in I,\, h \in H)$. This gives a **psychological game** in the sense of B&D (Section 6). An **assessment** is a profile of behavior strategies and beliefs $(\sigma_i, \beta_i)_{i\in I} \in \times_{i\in I}(\Sigma_i \times \Delta_i^2)$ such that $\Sigma_i = \times_{h\in H}\Delta\left(A_i(h)\right)$ and $\sigma_i$ is the plan $\alpha_{i,i}$ entailed by second-order belief $\beta_i$:

$$\sigma_i(a_i|h) = \alpha_{i,i}(a_i|h) = \beta_i\left(Z(h,a_i) \times \Delta_{-i}^1|h\right) \tag{12}$$

for all $i \in I$, $h \in H$, $a_i \in A_i(h)$. Eq. (12) implies that the behavior strategies contained in an assessment are implicitly determined by players' beliefs about paths; therefore, they could be dispensed with. Yet, we follow B&D and make behavior strategies explicit in assessments only to facilitate comparisons with the 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 $a = (a_j)_{j\in I} \in A(h)$,*

**(a)** $\alpha_i(a|h) = \prod_{j\in I}\sigma_j(a_j|h)$,

**(b)** $marg_{\Delta_{-i}^1}\beta_i(\cdot|h) = \delta_{\alpha_{-i}}$,

*where $\alpha_i$ is derived from $\beta_i$ and $\delta_{\alpha_{-i}}$ is the Dirac probability measure that assigns probability 1 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—conditional on each $h$—each $i$ expects each $j$ to behave in the continuation as specified by $j$'s plan $\sigma_j = \alpha_{j,j}$, even though $j$ has previously deviated from $\alpha_{j,j}$. All players thus have the same first-order beliefs. 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. Thus all players, essentially, have the same second-order beliefs (considering that they are introspective and therefore know their own first-order beliefs). These conditions faithfully capture the "trembling-hand" interpretation of deviations implicit in Kreps & Wilson's (1982) original definition of SE: if player $i$ observed deviations from the plans $\alpha_{-i}$ he ascribes to his co-players (according to his second-order beliefs $\beta_i$), instead of changing his mind about the co-players' plans, he would conclude that they made mistakes in carrying out their plans, but such mistakes are independent across nodes and the probability that they will occur in the future is negligible.

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

It can be checked that this definition is equivalent to the traditional one if players have standard preferences, *i.e.*, with a profile of utility functions $(v_i : Z \to \mathbb{R})_{i \in I}$ such that $u_i(h, a_i; \beta_i) = \mathbb{E}[v_i|(h, a_i); \alpha_i]$.[19] A special case is the **material-payoff game**, where $v_i = \pi_i$ for each $i \in I$.

**Remark 3** *Every material-payoff game with* perfect information *and* no relevant ties *has a unique SE, which is in pure strategies and can be computed by backward induction.*

As is known from previous work,[20] with psychological utilities uniqueness of equilibrium with deterministic plans may fail even in game forms with perfect information and with no relevant ties. The examples in Section 4.2 illustrate this for the case of anger-prone players. On the other hand, existence is guaranteed under mild conditions.

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

B&D prove a version of this existence result where first-order beliefs are modeled as belief systems over pure strategy profiles. Our setting adds personal histories, and here first-order beliefs are modeled as beliefs about paths. Given those modifications, the "trembling-hand" technique used in B&D's Theorem 9 can be applied to establish the general existence of psychological sequential equilibria in our framework. We omit the details.[21]

What we said so far about equilibrium does not assume specific functional forms. From now on, we focus on $u_i^{SA}$, $u_i^{ABB}$, and $u_i^{ABI}$. Since frustration and blame are continuous in beliefs, decision utility is also continuous, and we obtain existence in all cases of interest:

**Theorem 2** *Every game with SA, ABB, or ABI has at least one SE.*

## 4.2 Properties and examples

Next, we derive some general properties of how decision-makers with SA, ABB, or ABI behave. First, in pure-strategy SE's, players are never frustrated on the equilibrium path:

[19] According to the standard definition of SE, sequential rationality is given by global maximization over (continuation) strategies at each $h \in H$. By the One-Shot-Deviation principle, in the standard case this is equivalent to "local" maximization over actions at each $h \in H$.

[20] See Geanakoplos, Pearce & Stacchetti (1989) and B&D.

[21] A similar technique is used in the proof of Proposition 2 (first part) in the appendix.

**Proposition 1** *Let $(\sigma_i, \beta_i)_{i\in I}$ be an SE assessment of a game with SA, ABB, or ABI; if a history $h \in H$ has probability 1 under profile $(\sigma_i)_{i\in I}$, then*

$$\mathrm{F}_i(h';\alpha_i) = 0 \text{ and } \mathrm{Supp}\sigma_i(\cdot|h') \subseteq \arg\max_{a'_i\in A_i(h')} \mathbb{E}[\pi_i|\,(h', a'_i)\,;\alpha_i]$$

*for all $h' \preceq h$ and $i \in I$, where $\alpha_i$ is derived from $\beta_i$. Therefore, an SE strategy profile of a game with SA, ABB, or ABI with randomization (if any) only in the last stage is also a Nash equilibrium of the agent form of the corresponding material-payoff game.*

To illustrate, in Fig. B, $(f, n)$ can be an SE under ABB, and is a Nash equilibrium of the agent form with material-payoff utilities.[22] With (counterfactual) anger, $n$ becomes a credible threat. Proposition 1 also holds for the multistage extension of Section 5.

Recall that two assessments are **realization-equivalent** if the corresponding strategy profiles yield the same probability distribution over terminal histories:

**Proposition 2** *In every perfect-information (two-stage) game form with no chance moves and no relevant ties, the unique material-payoff equilibrium is realization-equivalent to an SE of the psychological game with ABI, ABB, or—with only two players—SA.*

The material-payoff SE of a perfect-information game must be in pure strategies (see Remark 3). By Proposition 1, players must maximize their material payoff on the path even if they are prone to anger. As for off-equilibrium path decision nodes, deviations from the material-payoff SE strategies can only be due to the desire to hurt the first-mover, which can only increase his incentive to stick to the material-payoff SE action.

It is quite easy to show by example that without perfect information, or with chance moves, a material-payoff SE need not be equivalent to an SE with frustration and anger. The same holds for some multistage game forms (cf. Section 5). Disregarding chance moves, randomization, and ties, the common feature of material-payoff SE that are not realization-equivalent to SE with frustration and anger is this (see Fig. A and Ex. 8 below): Start with a material-payoff SE and add anger to decision utility; now, at an off-path node after Ann deviates, frustrated Penny wants to hurt Bob, which implies rewarding Ann; this makes it impossible to incentivize both Ann not to deviate and Penny to punish Bob after Ann's deviation.

We next illustrate via examples how the SE concept works and how SA, ABB, and ABI may alter material incentives and produce different predictions. We begin with the "Hammering-one's-thumb" game which provides an example of displaced aggression: Under SA, Ann may take out her frustration on Bob if $\theta_a$ is sufficiently high, while ABB (both forms) and ABI preclude her from behaving aggressively towards Bob, as he is blameless.

[22] In the agent form of a game, each $h$ where player $i$ is active corresponds to a copy $(i, h)$ of $i$ with strategy set $A_i(h)$ and the same utility function as $i$.

**Example 7** Consider Fig. C. With $u_a^{ABB}$ (either version), or $u_a^{ABI}$, Andy will not blame Bob so his SE-choice is $N$. But with $u_a^{SA}$ Andy may choose $T$. Recall that $\mathrm{F}_a(B;\alpha_a) = 2(1-\varepsilon)+\varepsilon\alpha_a(N|B)-1$, so the more likely Andy believes it to be that he will take it out on Bob, the less he expects initially and the less frustrated he is after $B$. Yet, in SE, the higher is $\theta_a$ the more likely Andy is to take it out on Bob: Andy's utility from $N$ and $T$ is

$$u_a^{SA}(B,N;\alpha_a) = 1-\theta_a[2(1-\varepsilon)+\varepsilon\alpha_a(N|B)-1]\cdot 2,$$
$$u_a^{SA}(B,T;\alpha_a) = 0-\theta_a[2(1-\varepsilon)+\varepsilon\alpha_a(N|B)-1]\cdot 0 = 0.$$

Sequential rationality of SE implies that one possibility is $\alpha_a(N|B)=1$ and $u_a^{SA}(B,N;\alpha_a) \geq u_a^{SA}(B,T;\alpha_a)$, implying $\theta_a \leq \frac{1}{2(1-\varepsilon)}$. Another possibility is $\alpha_a(N|B)=0$ and $u_a^{SA}(B,N;\alpha_a) \leq u_a^{SA}(B,T;\alpha_a)$, implying $\theta_a \geq \frac{1}{2(1-2\varepsilon)}$. I.e., if Andy is sufficiently susceptible to SA, on bad days he takes his frustration out on Bob. If $\theta_a \in (\frac{1}{2(1-\varepsilon)}, \frac{1}{2(1-2\varepsilon)})$, we can solve for an SE where $u_a^{SA}(B,N;\alpha_a) = u_a^{SA}(B,T;\alpha_a)$ and $\alpha_a(N|B) = \frac{1}{2\varepsilon\theta_a} - \frac{1-2\varepsilon}{\varepsilon} \in (0,1)$.

The final case, where $\theta_a \in (\frac{1}{2(1-\varepsilon)}, \frac{1}{2(1-2\varepsilon)})$, illustrates how we cannot take for granted the existence of an SE in which players use deterministic plans (a point relevant also for $u_i^{ABB}$or $u_i^{ABI}$ in other games). Here this happens with a single active player, highlighting that we deal with a psychological game, as this could not be the case in a standard game. ▲

The next two examples highlight difference between ABB and ABI.

**Example 8** Consider Fig. A. Can material-payoff equilibrium outcome $(U,L)$ be part of an SE? The answer is yes under ABI and the blaming-unexpected-deviations version of ABB. To see this note that as first movers Ann and Bob act as-if selfish (as they are not frustrated; see Remarks 1 and 2). Hence, they would deviate if they could gain materially. In the SE, they would expect 5 if not deviating, making Ann the sole deviation candidate (she would get $6>5$ were Penny to choose $P$; for Bob, 5 is the best he can get). But Ann deviating can be dismissed since if $(D,L)$ were reached Penny would not blame Bob (her only punishable co-player) under either relevant blame function, so she would choose $N$ (regardless of $\theta_p$). Under SA and the could-have-been version of ABB, however, it may be impossible to sustain an SE with $(U,L)$; at $(D,L)$ Penny would blame each of Ann and Bob (as explained). By choosing $P$ she hurts Bob more than she helps Ann and would do so if

$$u_p^{ABB}((D,L),P;\alpha_p) > u_p^{ABB}((D,L),N;\alpha_p)$$
$$\Longleftrightarrow$$
$$0-6\theta_p\mathrm{B}_{pa}((D,L);\alpha_p) > 1-8\theta_p\mathrm{B}_{pa}((D,L);\alpha_p).$$

The rhs of the last inequality uses $\mathrm{B}_{pb}((D,L);\alpha_p) = \mathrm{B}_{pa}((D,L);\alpha_p)$. Since $\mathrm{B}_{pa}((D,L);\alpha_p) = \mathrm{F}_p((D,L);\alpha_p) = 1 > 0$, Penny would choose $P$ if $-6\theta_p > 1-8\theta_p \Longleftrightarrow \theta_p > 1/2$, so Ann would want to deviate and choose $D$. ▲

**Example 9** Consider Fig. B. By Proposition 2, every utility function discussed admits $(g, y)$ as an SE, regardless of anger sensitivity. To check this, just note that, if Bob expects $g$, he cannot be frustrated, so—when asked to play—he maximizes his material payoff. Under SA and ABB (both versions), $(f, n)$ qualifies as another SE if $\theta_b \geq 1/3$; following $g$, Bob would be frustrated and choose $n$, so Ann chooses $f$. Under ABI $(f, n)$ cannot be a SE. To verify, assume it were, so $\alpha_a(f) = 1$. Since the SE concept does not allow for players revising beliefs about beliefs, we get $\beta_b(\alpha_a(f) = 1|g) = 1$ and $\mathrm{B}_{ba}(g; \beta_b) = 0$; Bob maintains his belief that Ann planned to choose $f$, hence she intended to maximize Bob's payoff. Hence, Bob would choose $y$, contradicting that $(f, n)$ is an SE. Next, note that $(g, n)$ is not an SE under any concept: Given SE beliefs Bob would not be frustrated and hence choose $y$. The only way to observe rejected offers with positive probability in an SE is with non-deterministic plans. To find such an SE, note that we need $\alpha_a(g) \in (0, 1)$; if $\alpha_a(g) = 0$ Bob would not be reached and if $\alpha_a(g) = 1$ he would not be frustrated, and hence, he would choose $y$. Since Ann uses a non-degenerate plan she must be indifferent, so $\alpha_b(y) = 2/3$, implying that Bob is indifferent too. In SE, Bob's frustration is $\left[2\left(1 - \alpha_a(g)\right) + \frac{2}{3}\alpha_a(g) - 1\right]^+ = \left[1 - \frac{4}{3}\alpha_a(g)\right]^+$, which equals his blame of Ann under SA and ABB. Hence we get the indifference condition

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

$$\Longleftrightarrow$$

$$\alpha_a(g) = \frac{3}{4} - \frac{1}{4\theta_b},$$

where $\theta_b \geq 1/3$. The higher is $\theta_b$ the more likely Bob 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:

$$1 - \theta_b \mathrm{B}_{ba}(g; \beta_b) \cdot 3 = 0 - \theta_b \mathrm{B}_{ba}(g; \beta_b) \cdot 0$$

$$\Longleftrightarrow$$

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

The left term in braces is Bob's frustration, while

$$\frac{4}{3}\alpha_a(g) = 2 - \left[2(1 - \alpha_a(g)) + \frac{2}{3}\alpha_a(g)\right]$$

is the difference between the maximum payoff Ann could plan for Bob and the actual one. The first term is lower if $\alpha_a(g) \geq 3/8$; if, with that, we can solve the equation, we duplicate

the SA/ABB-solution; this is doable if $\theta_b > 1/3$. If $\theta_b \geq 2/3$, with ABI, there is a second non-degenerate equilibrium plan with $\alpha_a(g) \in (0, \frac{3}{8})$ where $\alpha_a(g) = 1/4\theta_b$; to see this, solve the ABI indifference condition assuming $\frac{4}{3}\alpha_a(g) \leq 1 - \frac{4}{3}\alpha_a(g)$. This SE exhibits starkly different comparative statics: The higher is $\theta_b$, the less likely Bob 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. ▲

The reason why $(f, n)$ in Example 9 cannot be an SE under ABI is that if Bob initially expects Ann to choose $f$, and she doesn't, so that Bob is frustrated, then he would rate her choice an unintended mistake and not blame her. We emphasize that this is due to assumptions that underlie the SE concept, *i.e.*, the "trembling-hand" interpretation of deviations, not to the formulation of ABI utility.

## 4.3 Threats and anger in leader-follower games

We now analyze the effect of frustration, anger, and blame in two-stage leader-follower game forms with perfect information, a class that includes the ultimatum game, the chain store game (Selten 1978), and the "pure threats game" of Klein & O'Flaherty (1993, Fig. 2). In such games there are two players, and only one is active in each stage. In the first stage the leader (denoted by $\ell$) is active. In the second stage, the follower (denoted by $f$) is active. Thus, the leader does not move in stage 2, the follower does not move in stage 1, and there is no third party. Formally:

**Definition 3** *A game form is called a* ***leader-follower*** *game if* $H = \{\varnothing\} \cup A(\varnothing)$, $I = \{\ell, f\}$, $I(\varnothing) = \{\ell\}$, *and* $I(a^1) = \{f\}$ *or* $I(a^1) = \emptyset$ *for every* $a^1 \in A(\varnothing)$.

Condition $H = \{\varnothing\} \cup A(\varnothing)$ of this definition says that no action of the leader terminates the game. This simplifies the exposition and is without loss of generality, because the follower's set of feasible actions may be a singleton after some $a^1$, in this case $I(a^1) = \emptyset$. For example, we interpret the Ultimatum Minigame of Fig. B as a game form where the responder is forced to accept the fair offer.

In leader-follower games both versions of ABB are behaviorally equivalent to SA. Formulae (9) and (10) each credit the full frustration on the first-mover of a leader-followers game, because each satisfies (8). Full blame for the follower's frustration is assigned to the leader. Let us write $u_{i,\theta_i}$ to make the dependence of $u_i$ on $\theta_i$ explicit; then (8) implies:

**Remark 4** *In leader-follower games, SA and ABB coincide, i.e.,* $u^{SA}_{i,\theta_i} = u^{ABB}_{i,\theta_i}$ *for all* $\theta_i$.

With more followers, or an inactive player, Remark 4 may not hold, as shown next:

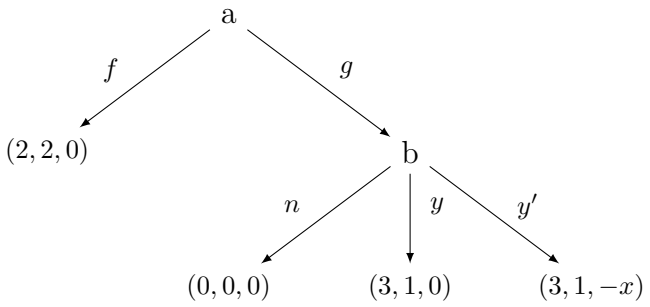


**Figure D.** Ultimatum Minigame with a bystander.

**Example 10** Consider the extension of the ultimatum minigame in Fig. D. The game adds a third player, Darryl (d, because we leave $c$ for chance), whose payoffs are represented by the third element in the payoff vectors. It also adds an alternative way for Bob to accept by choosing $y'$ which punishes Darryl by amount $x$. If Ann chooses $g$, assuming $\alpha_b(g) = \varepsilon$, we have

$$\mathrm{F}_b(g;\alpha_b) = (1-\varepsilon)\cdot 2 + \varepsilon\Big(\alpha_b(y|g) + \alpha_b(y'|g)\Big)\cdot 1 - 1 > 0.$$

With SA and $\theta_b > 0$, if $x > 3$, after $g$ Bob chooses $y'$ in another example of displaced aggression. But, since Darryl is passive and does not move in the first stage, with ABB Bob does not blame Darryl, and Bob is indifferent between $y$ and $y'$, regardless of $\theta_b$. ▲

In leader-follower games, our model captures an important aspect of the psychology of anger. Sell *et al.* (2009) argue that anger "is produced by a neurocognitive program engineered by natural selection to use bargaining tactics to resolve conflicts of interest in favor of the angry individual." We argue that our approach is consistent with this view. First note that our complete-information analysis assumes that the leader knows how anger-prone the follower is. This is a good approximation for interactions between agents who know each other well, and also a good approximation of face-to-face interactions, given that humans are good at reading facial cues to infer personality traits such as trait-anger (van Leeuwen *et al.* 2018). With this, our models predict that in leader-follower game forms anger-prone followers will obtain at least as large a material payoff as self-interested ones. In our evolutionary past, higher material payoffs meant longer survival and better access (for males) to sexual partners, both of which yield higher reproduction rates (e.g., Buss 2016). Hence, we argue that our approach supports the claim of Sell *et al.*

**Proposition 3** *In every leader-follower game form with no relevant ties, the expected material payoff of the follower in any SE with SA, or ABB, or ABI is at least as large as the material payoff of the follower in the unique material-payoff SE.*

To see why this is the case, consider an SE of the psychological game that yields a different expected material payoff to the follower than the material-payoff equilibrium. Since a first mover cannot be frustrated, all the actions chosen by the leader with positive probability in this SE must maximize his expected material payoff. Thus, if in this SE the leader deviates from the material-payoff equilibrium action, it must be the case that he correctly expects this action to be "punished" by the follower with a deviation from the follower's material-payoff maximizing reply. This in turn requires that the follower is frustrated by the material-payoff equilibrium action of the leader, hence, that his expected payoff in this SE is higher than in the material-payoff equilibrium.

In games forms with more than two followers, Proposition 3 need not hold. For example, there may exist equilibria where one follower behaves as-if self-interested, while the other is frustrated by the material-payoff equilibrium outcome. This can result in the first follower getting less than in the material-payoff equilibrium as shown by the following example.

**Example 11** Consider the game in Fig. *E.* The leader, Ann, chooses between Left ($l$) and Right ($r$). In the left subgame Bob chooses between the payoff profile $(7, 1, 3)$ and 0 for all players. In the right subgame Darryl chooses between the payoff profile $(6, 2, 2)$ and 0 for all players. The material-payoff SE is $(l, y, y')$ and in this equilibrium Darryl gets 3. However, with SA, ABB, or ABI, for sufficiently large values of $\theta_b$ and for all $\theta_d \geq 0$, the strategy profile $(r, n, y')$ can also be an SE. To see this note that if Ann deviates to $l$, Bob will be frustrated, and if $\theta_b$ is large enough Bob will choose $n$ after $l$. In this equilibrium Darryl gets 2, a payoff smaller than in the material payoff equilibrium. ▲

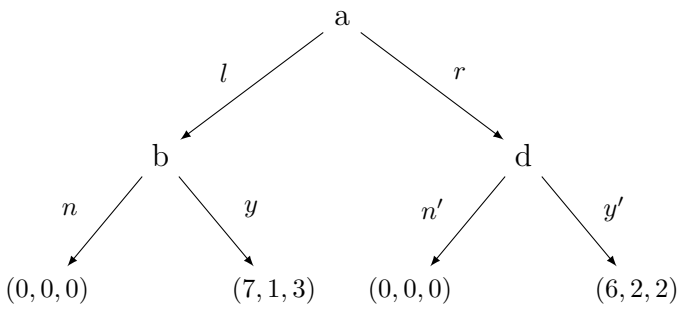


**Figure E.** A game with two followers.

Next, we demonstrate that our models of anger are behaviorally relevant only in the subclass of leader-follower games involving threats. As above, consider a leader-follower game form with no relevant ties. In these games the leader has a unique best response to each pure

strategy $s_f \in \times_{a_\ell \in A_\ell(\varnothing)} A_f(a_\ell)$ of the follower. We let

$$r_\ell(s_f) = \arg \max_{a_\ell \in A_\ell(\varnothing)} \pi_\ell(a_\ell, s_f(a_\ell))$$

denote this best response. Let $(\bar{a}_\ell, \bar{s}_f) = (r_\ell(\bar{s}_f), \bar{s}_f)$ denote the unique material-payoff SE, where $\bar{s}_f$ is the material equilibrium pure strategy of the follower. Note also that, by the no-relevent-ties assumption, for every pure strategy of the follower, the leader has a unique best response. With this, we can define a threat as a strategy of the follower that penalizes the leader for playing the material payoff SE strategy:

**Definition 4** *In any leader-follower game form with* no relevant ties*, a* ***threat*** *of player $f$ is a strategy $\hat{s}_f$ such that*

1. *A deviation to the threat from the material-payoff SE strategy by the follower harms the leader: $\pi_\ell(\bar{a}_\ell, \hat{s}_f(\bar{a}_\ell)) < \pi_\ell(\bar{a}_\ell, \bar{s}_f(\bar{a}_\ell))$.*
2. *The best response to the threat benefits the follower compared to the material-payoff SE: let $\hat{a}_\ell = r_\ell(\hat{s}_f)$, then $\pi_f(\hat{a}_\ell, \hat{s}_f(\hat{a}_\ell)) = \pi_f(\hat{a}_\ell, \bar{s}_f(\hat{a}_\ell)) > \pi_f(\bar{a}_\ell, \bar{s}_f(\bar{a}_\ell))$.*

Our definition of a threat is inspired by but differs from that of Klein & O'Flaherty, who focus on multi-stage games with pure-strategy material-payoff equilibria. Note that the threat must be costly to implement along the path of the material payoff SE, because — by condition 1 — it differs from the unique material best response. Thus, the above conditions incorporate the notion of a threat that would not be credible if the follower were known to be a material payoff maximizer. The paradigmatic example of a game form with threats is the Ultimatum Game: indeed, the strategy of rejecting the greedy offer in the Ultimatum Minigame of Figure B is a threat.

We can use the conditions in Definition 4 to (partially) characterize the SE behavior of anger-prone followers:

**Proposition 4** *In every leader-follower game form with no relevant ties where the follower's has a* threat*, there exists an SE of the psychological game with SA/ABB (when the anger sensitivity parameter is sufficiently large) such that*

1. *The follower's strategy is a threat.*
2. *The leader does not play his material-payoff SE action.*
3. *The follower's material payoff is strictly greater than in the material-payoff SE.*

The proof of Proposition 4 (see the Appendix) involves demonstrating that the consistent assessment where the follower plays a given threat (and the leader best responds to the threat) is sequentially rational when the follower is sufficiently prone to anger, and therefore it constitutes an SE with SA and ABB. This construction does not work with ABI for the reasons explained in Example 9 about the Ultimatum Game: under the "trembling-hand" interpretation of deviations inherent in the SE concept, a deviation from the action $\hat{a}_\ell$ that the follower wants to induce is interpreted by him as unintentional, so that he does not blame the leader and the threat is not credible.

The next proposition shows that, if the game form does not include a threat, then the follower behaves as-if self-interested in every pure SE:

**Proposition 5** *In every leader-follower game form with* no relevant ties *and* without threats*, every pure SE of the psychological game with SA/ABB or ABI is realization equivalent to the material-payoff SE.*

To illustrate, anger is not relevant in the trust game of Berg *et al.* (1995) (and the mini-trust game in B&D's Fig. 1). In general, behavioral patterns that require players to place positive weight on a co-player's material payoff cannot be explained via anger.

# 5 Multistage extension

In a multistage game form, a (nonempty) nonterminal history is a sequence of action profiles, $h = (a^1, ..., a^t)$ where $t \geq 1$. As in the two-stage case, we assume that actions are observable; hence, every non-terminal history is public. Our notation for the multistage setting is essentially the same as before. The set of sequences observable by player $i$ also includes personal histories of the form $(h, a_i)$: $H_i = H \cup \{(h, a_i) : h \in H, a_i \in A_i(h)\}$.

A belief system for $i$ over paths and beliefs of others is an array of probability measures $\beta_i = (\beta_i(\cdot|h_i))_{h_i \in H_i}$ satisfying (3) and (4), which apply to the multistage setting as well. Also the notation on beliefs is as before: $\alpha_i \in \Delta_i^1$, $\beta_i \in \Delta_i^2$, and $\alpha_i$ is the first-order belief system derived from $\beta_i$ when they appear in the same formula. Therefore, the definition of SE from Section 4 can be applied without modifications, and Theorem 2 implies the existence of SE with SA, ABB, or ABI in the multistage setting.

We distinguish two extreme scenarios according to the behaviorally relevant periodization: In the **slow-play** scenario, stages correspond to periods, and player $i$'s reference expectation to determine frustration at the beginning of period (stage) $t+1$ is given by his belief at the beginning of period $t$. In the **fast-play** scenario, the game's different stages occur in the same period and the relevant reference expectation of $i$ in stage $t$ is given by his initial belief (at

the root).[23] In either case, we maintain the assumption that blame is continuous in beliefs, capped by frustration, and equal to frustration in the case of SA.

## 5.1 Slow play

We start with this scenario because it allows for a relatively simple extension of the two-stage setting, with initial beliefs replaced by one-period-lagged beliefs: For any non-terminal history of the form $h = (\bar{h}, a)$ the frustration of $i$ conditional on $h$ given $\alpha_i$ is

$$\mathrm{F}_i\left(h;\alpha_i\right) = \left[\mathbb{E}[\pi_i|\bar{h};\alpha_i] - \max_{a_i \in A_i(h)} \mathbb{E}[\pi_i|\left(h,a_i\right);\alpha_i]\right]^+.$$

(When $\bar{h} = \varnothing$ and $h = a^1$, we are back to the two-period formula.) The decision utility of action $a_i \in A_i(h)$ has the general form (5), where the blame functions $\mathrm{B}_{ij}\left(h;\beta_i\right)$ are of the SA, ABB, or ABI type. Specifically: $\mathrm{B}_{ij}\left(h;\beta_i\right) = \mathrm{F}_i\left(h;\alpha_i\right)$ for SA, whereas the could-have-been blame, blaming deviations, and blaming intentions can be defined with straightforward adaptations of (9), (10), and (11) respectively; therefore we omit the details.

This extension of the two-stage setting has the stark feature that past frustrations do not affect current behavior. (A more nuanced version of the model might feature a decaying effect of past frustrations.)

A detail in modeling game forms becomes relevant in the slow play scenario: We have to explicitly allow for non-terminal histories after which no player (not even chance) is active, such as history $g$ in Fig. F. At such histories there is only one feasible action profile, as each player has only one feasible action, to wait. In the two-periods setting this detail is irrelevant: If nobody is active at the root, play effectively starts (and ends) in the second period; if nobody is active at $a^1$, it can be modeled as a terminal history. With more than two periods, having to wait may affect behavior.

**Example 12** Consider Fig. F. Suppose that Bob initially expects $f$ with positive probability. In period 2, after $g$, he is frustrated; however he cannot hurt Ann immediately because he has to wait. In period 3, Bob's lagged expectation has fully adapted, and he is not frustrated. According to our slow-play model, the frustration experienced by Bob in period 2 does not affect his decision utility in period 3: Bob "cools off" and behaves as-if selfish.[24] Therefore the unique SE of the game is $(g, w, y)$, where $w$ denotes waiting. ▲

[23] Applications may involve intermediate cases, as in alternating-offer bargaining models where a period comprises two stages. The two extremes convey the main ideas.

[24] See e.g. Grimm & Mengel (2011) and Oechssler *et al.* (2015) for experiments.

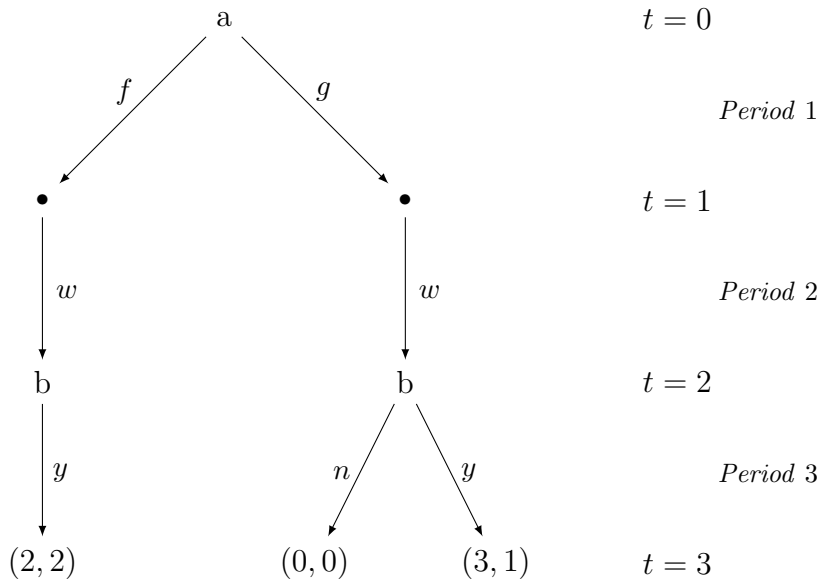


**Figure F.** Ultimatum Minigame with delayed reply.

## 5.2 Fast play

All stages now belong to the same period, and $i$'s frustration at $h$ given $\alpha_i$ is calculated with reference to initial expectations, as in Section 3:

$$\mathrm{F}_i(h;\alpha_i) = \left[\mathbb{E}[\pi_i;\alpha_i] - \max_{a_i \in A_i(h)} \mathbb{E}[\pi_i|(h,a_i);\alpha_i]\right]^+.$$

This implies that *there cannot be any cooling off* due to reference-point acclimatization. Formally, histories where nobody (not even $c$) is active play no role and therefore can be deleted from the game form without affecting the analysis, unlike with slow play. For example, in the fast-play scenario, game form of Fig. F is equivalent to the one of Fig. B.

With this, the fast-play frustration formula can be plugged into the SA decision utility function (7). As for the ABB decision utility, property (8) of $\mathrm{B}_{ij}$ extends to the multistage setting as follows:[25]

$$\mathrm{B}_{ij}(h;\alpha_i) = \begin{cases} 0, & \text{if } j \notin I(h') \text{ for all } h' \prec h, \\ \mathrm{F}_i(a^1;\alpha_i), & \text{if } \{j\} = I(h') \text{ for all } h' \prec h. \end{cases} \tag{13}$$

[25]Recall that (8) is a necessary condition that blame must satisfy on top continuity in beliefs (which guarantees existence of equilibria), it is not a full specification of the blame function for all game forms. This holds *a fortiori* for the extension given here.

In words, $i$ cannot blame $j$ if $j$ was never active in the past, and $j$ is fully blamed if he was the only active player. An extension of could-have-been blame satisfies this property:

$$\mathrm{B}_{ij}(h;\alpha_i) = \min\left\{\left[\max_{h'\prec h, a'_j\in A_j(h')} \mathbb{E}\left[\pi_i|(h',a'_j);\alpha_i\right] - \mathbb{E}[\pi_i|h;\alpha_i]\right]^+, \mathrm{F}_i(h;\alpha_i)\right\}. \qquad (14)$$

We can follow a similar logic to extend ABI.

**Proposition 6** *If* $\mathrm{B}_{ij}$ *is defined by (14), then it satisfies (13).*

We now illustrate our definition, elucidating a modeling choice:

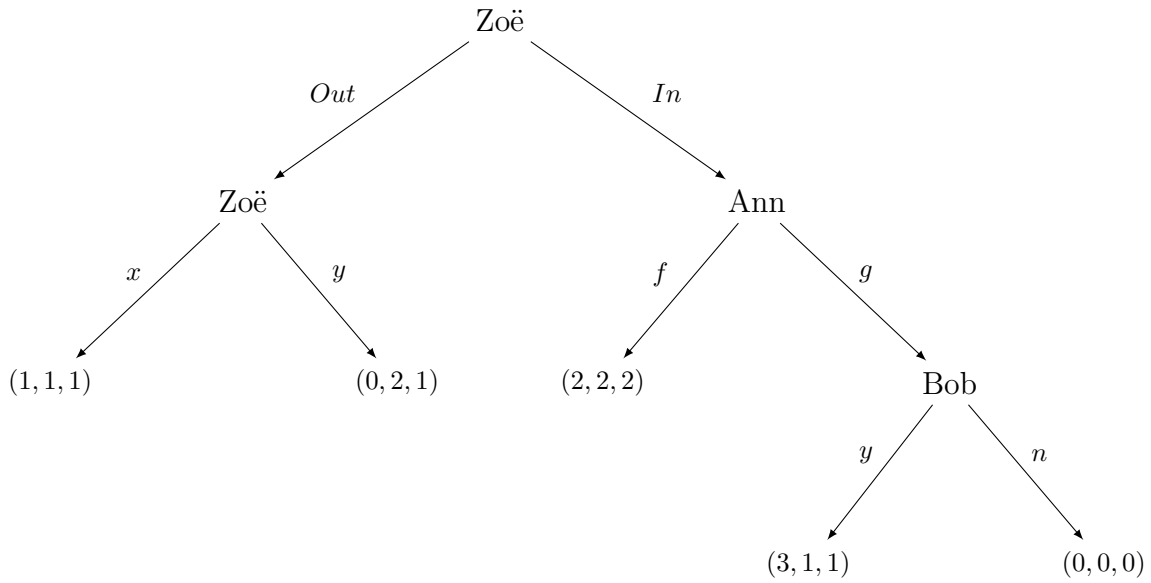


**Figure G.** Multistage Ultimatum featuring Zoë.

**Example 13** Consider the game form in Fig. G (material payoffs are in alphabetical order). If Zoë chooses $In$, then Ann and Bob interact in an ultimatum minigame, but Zoë may instead exercise outside options and play $(Out, x)$ or $(Out, y)$. Zoë's payoffs equal Bob's, except following $(Out, y)$ where a payoff transfer from Ann to Bob occurs, relative to $(Out, x)$. Can strategy profile $(In\text{-}x, f, n)$ be an SE under ABB? Given equilibrium beliefs, this is the case if $0-\theta_b\cdot 1\cdot 0 \geq 1-\theta_b\cdot 1\cdot 3$, or $\theta_b \geq 1/3$. The calculation involves Bob blaming Ann, not Bob blaming Zoë, because if Zoë switched from $In$ to $Out$ (thus implementing $(Out, x)$ instead of $In$) this would not improve Bob's payoff. This reflects a non-obvious modeling choice: Our

definition assesses blame on the basis of single-agent deviations from the realized path, but if Bob alternatively assessed blame on the basis of multi-agent deviations, including off-realized-path deviations, he would consider that Zoë could have played $(Out, y)$. She would then have increased Bob's payoff from 1 to 2, preventing his frustration of 1. If Bob's blame of Zoë were thus 1, then $(In\text{-}x, f, n)$ would be an SE under ABB if $0 - \theta_b \cdot 1 \cdot 0 \geq 1 - \theta_b \cdot 1 \cdot 3 - \theta_b \cdot 1 \cdot 1$, or $\theta_b \geq 1/4 \neq 1/3$. (This also shows that SE under ABB is not invariant with respect to coalescing sequential moves.) Finally, note that also $(In\text{-}y, f, n)$ is an SE under ABB in the fast-play scenario for $\theta_b \geq 1/4$, because at $(In, g)$ Zoë would be blamed for not switching to $Out$ (implementing $(Out, y)$); but it is an SE under ABB in the slow-play scenario for larger parameter values, $\theta_b \geq 1/3$, because Bob would be frustrated only in the third period, after $(In, g)$, and Zoë—who played in the first—could not be blamed. ▲

The single- *vs.* multi-agent deviation issue illustrated here can arise also in two-stage games (with simultaneous moves), but the point is clearer, and perhaps more relevant, with more than two stages. In the example, we described our favored formulation as a "non-obvious modeling choice," and we close this section by thrice defending it: It harmonizes well with how we define rational play, where players optimize only locally (although in SE they predict correctly and choose as planned). The (hinted at) alternative definition would be formally convoluted. It is an open issue which formulation is empirically more relevant; for now, we stick with what is simpler.

## 5.3 Counterfactual anger and unique SE in hold-up

Anger, and in fact emotions more generally, can shape behavior without occurring. If anger is anticipated, this may steer behavior down alternative paths (cf. Proposition 1). We already saw examples, *e.g.*, $(f, n)$ may be an SE in the Ultimatum Minigame, alongside $(g, y)$. Our next example highlights how there may be circumstances where the SE is *unique* and has that property. It also illustrates a difference between fast and slow play.

**Example 14** Modify the Ultimatum Minigame by adding an initial move for Bob, as in Fig. H, to get an illustration of a hold-up problem (cf. Dufwenberg, Smith & Van Essen 2013).[26] Under fast play, for each utility function seen so far,[27] if $\theta_b > 2/3$, there is a unique SE: Bob uses plan $(r\text{-}n)$, Ann plans for $f$. To verify this, the key step is to check that if Bob plans for $(\ell, y)$ and Ann for $g$ this is *not* an SE; if Bob initially expects \$1.5, off-path at $(r, g)$, he would be frustrated and deviate to $n$. ▲

[26] Bob and Ann face a joint business opportunity worth $(2, 2)$ via path $(r, f)$; however, $r$ involves partnership-specific investment by Bob, which Ann can exploit choosing $g$ (reneging), etc. As always, we list payoffs by alphabetical order of players: $(\pi_a, \pi_b)$.

[27] Except the blaming-unexpected-deviations version of ABB, which we did not define explicitly for fast play.

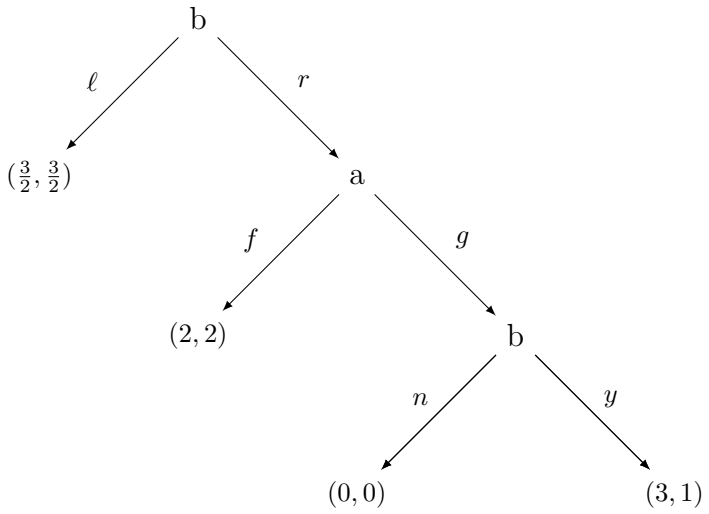


**Figure H.** Hold-up.

With slow play, by contrast, with $\theta_b > 2/3$, there are multiple SE, exactly as in the Ultimatum Minigame. In particular, both $(r\text{-}n, f)$ and $(\ell\text{-}y, g)$ are SE; in the latter, Bob's updated expected payoff after (counterfactual) action $r$ is only \$1, hence he cannot be frustrated by $g$.

# 6 Comparison with other models

We shall now compare and contrast the behavioral predictions resulting from frustration and anger with other models of strategic and social behavior. We first consider distributional preferences, which transform material payoffs at terminal histories but which otherwise retain the standard assumption that choices depend solely on their consequences in terms of material payoffs. We then discuss models of reciprocity, and some alternative approaches to modeling anger.

## 6.1 Distributional preferences

Like anger, models of distributional preferences such as inequality aversion (e.g. Fehr & Schmidt 1999, Bolton & Ockenfels 2000) predict costly punishment. However, a number of studies demonstrate that the decision to engage in costly punishment depends upon both payoffs reached at terminal histories as well as payoffs from unreached histories.[28] Our approach

[28]See e.g. Brandts & Solá (2001), Charness & Rabin (2002), Nelson (2002), Falk *et al.* (2003, 2008), Sutter (2007).

can captures this non-consequentialist aspect of behavior, while distributional preferences cannot. For example, in the ultimatum minigame of Fig. B, Bob's decision to reject the greedy offer may depend upon not only the payoffs from accepting or rejecting the offer, but also upon the payoff that Ann could have given Bob had she chosen the fair offer. Holding payoffs and other beliefs constant, our models predict that Bob will be more likely to reject the greedy offer when either he assigns higher probability to receiving the fair one, or when the fair payoff is increased. In general, distributional preferences cannot capture behavioral patterns which depart from consequentialism, while our models can.

The models in this paper assume that players care only about material payoffs and anger, disregarding distributional considerations. We do so to highlight the effects of frustration, blame, and anger on behavior in strategic interaction. However, real-world decision makers may have a mixture of material, distributional, and psychological motivations (e.g. Falk & Fischbacher 2006). Although we do not pursue the modeling of distributional concerns, doing so might involve straightforward modifications of our material payoff function. We leave this for future work.

## 6.2 Reciprocity

Negative reciprocity à la Rabin (1993), Dufwenberg & Kirchsteiger (2004), and Falk & Fischbacher (2006) joins anger as a motivation that can trigger hostility. Also like anger, negative reciprocity can result in non-consequential behavior, but anger and negative reciprocity differ in key ways. The following sketched comparison is with Dufwenberg & Kirchsteiger's notion of sequential reciprocity equilibrium (SRE; refer to their article for formal definitions).

In the Hammering one's thumb game (Fig. C), Andy may take it out on Bob if he is motivated by simple anger. Were he motivated by reciprocity, this could never happen: Bob's kindness, since he is a dummy-player, equals 0, implying that Andy chooses as-if selfish. Reciprocity here reflects intuitions similar to the ABI concept, but that analogy only carries so far, as we show next.

Reciprocity also allows for "miserable equilibria," where a player reciprocates expected unkindness before it occurs. For example, in the Ultimatum Minigame of Fig. B, $(g, n)$ may be a SRE. Ann makes offer $g$ despite believing that Bob will reject; given her beliefs about Bob's beliefs, Ann perceives Bob as seeing this coming, which makes him unkind, so she punishes by choosing $g$. Such self-fulfilling prophecies of destructive behavior have no counterpart under any of our anger notions. Since Ann moves at the root, Remarks 1 and 2 demonstrate that she cannot be frustrated, and hence chooses as-if selfish.[29] By Propostion

[29] Another example is the hold-up game of Fig. H. We gave conditions where $(r\text{-}n, f)$ was the unique SE. If Ann and Bob were reciprocal, $(\ell\text{-}n, g)$ and $(r\text{-}n, g)$ could be SRE, with miserable interaction, respectively, off and on the equilibrium path.

1, sacrificing material payoff to harm a co-player never occurs on the path of a pure-strategy SE with ABB, ABI, or (in two-player games) SA.

So under our definition, anger is "hot": in pure-strategy SE's, angry players do not expect to sacrifice to harm others. In addition, the cooling-off effects discussed earlier (Section 5) have no counterpart in reciprocity theory, which makes the same prediction in Figs. B and F. Reciprocal players do not cool off. "*La vengeance est un plat qui se mange froid.*"

## 6.3 Other models of anger

**Card & Dahl (C&D) and Kőszegi & Rabin (K&R)** C&D show that reports of domestic abuse go up when football home teams favored to win lose. They argue that this is in line with K&R's (2006, 2007) theory of expectations-dependent reference points. K&R model the loss felt when a player gets less than he expected, which one may think of as a form of disappointment with negative valence (cf. Bell 1985, Loomes & Sugden 1986). However, K&R do not model other-regarding preferences directly: they focus on the consequences of their model for individual decisions. Our models study the social consequences of frustration: frustration results in lower weights on coplayer payoffs, and hence encourages costly punishment. Our simple anger model and the example of hammering-one's thumb encapsulates C&D's result.

A key difference between this paper and K&R is that in their work anticipation of the negative valence of future frustrations influences decision utility. Our decision makers are influenced by past frustrations, rather than future ones. Important modeling details then distinguish how we define frustration and how K&R define loss (*e.g.*, how we cap frustration using the highest attainable payoff).

**Rotemberg** 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 angry and punish a co-player whom they come to believe has an altruism parameter lower than some 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 some cutoff value. If a person acts so that $i$ is able to reject this hypothesis, individual $i$ gains ill-will towards this person" (Rotemberg 2008, p. 464).

Rotemberg shows how his model impressively captures the action in his data sets. It is natural to wonder whether our approach, which is structured very differently from his (e.g., we make no reference to altruism), could achieve that too. As regards behavior in ultimatum (and some other) games, there is already some existing evidence that is consistent with our approach; see the discussion regarding experiments below. Regarding pricing, we leave for

empirical economists the task of exploring the topic.

**Winter, Brams** Winter (2014) and Winter *et al.* (2016) model anger and other emotions in games with a version of the indirect evolutionary approach:[30] Like us, Winter *et al* assume that preferences over outcomes are "emotional" and endogenous, but we differ in the way we model emotions and make them endogenous. We assume that emotions depend on endogenous beliefs, while Winter *et al* model the rest points of an adaptation process of belief-independent preferences. Brams (2011) studies anger in sequential interactions by modeling players who take turns changing the state of a $2 \times 2$ payoff matrix and receive payoffs at the end of the game. However, like Winter's, his model of anger is independent of beliefs, while we argue that beliefs are central to emotions.

**Akerlof** Akerlof (2016) models anger in a 2-person Bayesian game where the first mover decides whether or not to follow a rule, and the second mover decides whether or not to punish the first mover. In the model, Player 1's compliance with the rule is motivated by a sense of duty. Player 2 may be sensitive to anger from noncompliance if she thinks that a "reasonable person," modeled as a person with similar preferences to Player 2, would comply with the rule. Similarly to Rotemberg, Akerlof motivates costly punishment via preferences over others' types. In contrast, we assume that anger and aggression arise from frustration and payoff expectations. Both Rotemberg's and Akerlof's approach begin with norms about behavior, and condition anger upon violation of those norms. In contrast, we develop models that reflect the psychology of frustration and anger.

# 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 interaction, and experimental or other evidence will help to assess empirical relevance and suggest revised formulas. We conclude by discussing sundry topics that may help gain perspective on, build on, or further develop our work.

**Frustration** Consider substituting $\mathbb{E}[\pi_i; \alpha_i] - \mathbb{E}[\pi_i | a^1; \alpha_i]$ for $\mathrm{F}_i(a^1; \alpha_i)$ of Section 3. This alternative would measure $i$'s *actual* diminished expectations at $a^1$, unlike $\mathrm{F}_i(a^1; \alpha_i)$ which reflects diminished expectations relative to what $i$ believes is the most he can get (which we think of as the adequate way to represent the unexpected unavailability of something cared

[30] See, for example, Güth & Kliemt (1988).

about). To appreciate how dramatically this change would impact behavior, consider a two-player common-interest game: Ann chooses *Out* or *In*; in the former case the game ends with payoffs $(1,1)$, in the latter case Bob chooses between $(0,0)$ and $(2,2)$. *Mutatis mutandis*, for high enough $\theta_b$, with the alternative, under SA and ABB, there is an SE where Ann chooses *Out* and Bob would go for $(0,0)$. Following *In*, Bob would be frustrated because he (so-to-say) sees himself as locked-in with his stage-2 planned action. Our formulation of $\mathrm{F}_i(a^1;\alpha_i)$ rules that out.

Take a binary gamble where with probability $p>0$ Ann wins $\$x>0$, and otherwise gets \$0. Her frustration, using our definition, equals her initial expectation: $p\cdot x$. This embodies strong implications for how frustrations compare across contexts, *e.g.* the frustration of a highly expected failure to win the state lottery versus that of some unlikely small loss. We are agnostic as regards empirical relevance, but alert the reader to the issue.[31]

The psychological evidence (cited in Section 1) says a player becomes frustrated when his goals are unexpectedly thwarted. We addressed but one aspect: own material rewards. Cases 1-3 indicate broad applied potential. Yet our focus is restrictive, as one may imagine other sources of frustration:

> **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 Tea Party/Occupy-Wall Street movements?

In case 4, an early adopter is frustrated because he regrets he already bought, not because new information implies his expected rewards drop. In case 5, even an activist who is materially unaffected personally may be frustrated because of unexpected perceived unfairness. These examples are not exhaustive; further sources of frustration may *e.g.* involve shocks to self-esteem.[32] Techniques analogous to those we have developed may be applicable in these cases, but going in these directions is left for future research.

As regards the effects of frustration, we considered changes to a player's utility but neglected other plausible adjustments. Gneezy & Imas report data from an intriguing experiment involving two-player zero-sum material payoff games. In one game players gain if they are strong, in another if they are smart. Before play starts, one subject may anger his opponent

[31] The example involves one-player with a dummy-choice only to facilitate the frustration-calculation; interesting testable implications obviously arise more generally, *e.g.* in modified versions of the hammering one's thumb game.

[32] See Baumeister *et al.* (1996) for an interesting discussion linking (threatened) self-esteem and violence.

and force him to stay in the lab to do boring tasks. A thus frustrated player's performance is enhanced when strength is beneficial (possibly from increased adrenaline flow), but reduced when cool logic is needed (as if an angered player becomes cognitively impaired). Our approach can capture the first aspect, but not the second: We can let consequences of actions depend also on beliefs (*e.g.*, because emotions affect strength or speed; cf. Rauh & Seccia 2006); this ultimately translates into belief-dependent utility (or cost) of actions. However, to model the second effect, we would need a theory of endogenous cognitive abilities.

**Valence and action-tendency** Psychologists classify emotions in multiple ways. Two prominent aspects are **valence**, the intrinsic pleasantness or aversiveness of an emotion, and **action-tendency**, or how behavior is shaped as the emotion occurs. Both notions have bearing on anger. For example, most psychologists believe anger has negative valence (see, *e.g.*, Harmon-Jones & Sigelman 2001, p. 978). 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 stress concerns its action-tendency of aggression, not its valence. In developing our theory, we have exaggerated this, abstracting away from frustration avoidance, while emphasizing frustration-induced aggression. This is reflected in the decision utility functions, which are shaped by current frustration, but not by the anticipation of the negative valence of future frustrations.[33]

**Blame** We explored various ways a player may blame others, but other notions are conceivable. For example, with anger from blaming behavior $i$'s blame of $j$ depends on what $i$ believes he would truly get at counterfactual histories, rather than the most he could get there. We view this modeling choice as reflecting local agency; $i$'s current agent views other agents of $i$ 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, not 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 notion.

Our blame concepts one way or another assess the *marginal* impact of others. For example, consider a game where $i$ exits a building while all $j \in I \backslash \{i\}$, unexpectedly to $i$, simultaneously hurl buckets of water at $i$, who gets soaked. According to our approach, $i$ cannot blame any $j$

[33] In previous work we modeled another emotion: guilt (*e.g.*, Battigalli & Dufwenberg 2007, Chang *et al.* 2011). To gain perspective note how our approach to anticipation of valence and action-tendency was then reversed. Guilt has valence (negative!) as well as action-tendency (say to engage in "repair behavior"; see, *e.g.*, Silfver 2007). In modeling guilt we highlighted anticipation of its negative valence while neglecting action-tendency.

as long as there are at least two hurlers. One could imagine that $i$ alternatively blames, say, all the hurlers on the grounds that they *collectively* could thwart $i$'s misery, or that $i$ splits the blame among all hurlers. Halpern (2016, Chapter 6) explores such issues and develops a model that assigns positive blame even when outcomes are overdetermined.

People may also blame others in unfair ways, *e.g.* nominating scapegoats. Our notions of SA and ABB may embody such notions to some degree. However, it has not been our intention to address this issue systematically.

Several recent experiments explore interesting aspects of blame (Bartling & Fischbacher 2012, Gurdal *et al.* 2014, Celen *et al.* 2017). We emphasize that our focus on blame is restricted to its relation to frustration, not on reasons besides frustration that may lead people to blame each other.[34]

**Anger management** People aware of their inclination to be angry may attempt to manage or contain their anger. Our players anticipate how frustrations shape behavior, and they may avoid or seek certain subgames because of that. However, there are interesting related phenomena we do not address: Can $i$ somehow adjust $\theta_i$ say by taking an "anger management class?" If so, would rational individuals want to raise, or to lower, their $\theta_i$? How might that depend on the game forms they play? These are potentially relevant questions related to how we have modeled action-tendency. Further issues would arise if we were to consider aspects involving anticipated negative valence of future frustrations, or bursts of anger.

**Experimental testing** Our models tell stories of what happens when anger prone players interact. It is natural to wonder about empirical relevance. Experiments may shed light.

A few studies that measure beliefs, emotions, and behavior together provide support for the notion that anger and costly punishment result from outcomes which do not meet expectations. Pillutla & Murnighan (1996) find that reported anger predicted rejections better than perceived unfairness in ultimatum games. Fehr & Gächter (2002) elicit 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.

Other studies connect unmet expectations and costly punishment in ultimatum games. Falk *et al.* (2003) measure beliefs and behavior in ultimatum minigames; higher proportions of rejections of disadvantageous offers when responders' expected payoffs are higher, consistent

[34]For example, Celen *et al.* (2014) present a model where $i$ asks how he would have behaved had he been in $j$'s position and had $j$'s beliefs. Then $i$ blames $j$ if $j$ appears to be less generous to $i$ than $i$ would have been, and may blame $j$ even if $i$ is not surprised/frustrated. Or imagine a model where players blame those considered unkind, as defined in reciprocity theory, independently of frustration.

with our models.[35] Schotter & Sopher (2007) measure second-mover expectations, 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 emotions and expectations in power-to-take games (which resemble ultimatum games, but allows for partial rejections).[36] Second-mover expectations about first-mover "take rates" are key in the decision to destroy income, and anger-like emotions are triggered by the difference between expected and actual take rates. The difference between actual and reported "fair" take rate is not significant in determining anger, suggesting that deviations from expectations, rather than from fairness benchmarks, drive anger and the destruction of endowments.

Apropos the cooling off effects discussed in Section 5, Grimm & Mengel (2011) run ultimatum games that force some responders to wait ten minutes before making their choice. Without delay, less than 20% of low offers were accepted while 60–80% were accepted if the acceptance decision were delayed.

A literature in neuroscience connects expectations with 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 experimenter controls the distribution of offers in order to manipulate beliefs. Rejections occur more often when subjects expect higher offers. The authors connect norm violations (i.e., lower than expected offers) with reward prediction errors from reinforcement learning, which are known to be the computations instantiated by the dopaminergic reward system. Xiang *et al.* note that "when the expectation (norm) is violated, these error signals serve as control signals to guide choices. They may also serve as the progenitor of subjective feelings."

It would be 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 answer depend on context (*e.g.*, is SA 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 seems to indicate some 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 not measured).

A few recent studies are directly motivated by our work. Persson (2018) presents a test of simple anger. He explores the Hammering one's thumb game, and documents that frustrations

[35] See the data in Fig. 2 and Table 1 of their paper. Brandts & Solá (2001) find similar behavioral results (see Table 1). Compare also with our discussion of Example 3 in Section 3.2.

[36] Bosman & van Winden (2002), Bosman *et al.* (2005), Reuben & van Winden (2008).

occur much as predicted and yet no punishments occur. His results thus favor ABB or ABI over SA in that context. More recently, we have ourselves begun to study our models in the laboratory: Aina *et al.* (2018) devise tests that manipulate the responder's payoff from the proposer's outside option in mini-ultimatum games, and Dufwenberg *et al.* (2018*a*,*b*) test predictions that link anger to verbal promises & threats; in each case supporting evidence is reported.

Our models are abstractions. We theorize 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 emotions like guilt, disappointment, 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 (in addition to concern for material payoffs). This statement may seem trivially obvious, but it has subtle implications for how to evaluate experimental work. Experimental data may find that one of the forms of motivation that our theory abstracts away from affects subjects' choices. However, to reject our theory, it may be more relevant to ask if behaviors that are in fact driven by anger (as measured by, *e.g.*, emotion self-reports, physiological activity, or both, as in Chang *et al.* 2011) diverge from our predictions. If they were, that might indicate that our theory could benefit from revision.

**Applications** Formulating, motivating, and elucidating the key definitions of our models is more than a mouthful, so we have not taken this paper in the direction of doing applied economics. Make no mistake about it though, the hope that our models will prove useful for such work has been a primary driving force. Our psychologically grounded models of frustration, anger, and blame may shed light on many of the themes (e.g. pricing, violence, politics, recessions, haggling, terror, and traffic) that we listed at the start of this paper. We hope to do some work in these directions ourselves.

# A Appendix

This appendix contains proofs of the results stated in the main text.

## A.1 Preliminaries

To ease exposition, some of the key definitions and equations contained in the main text are repeated below.

For each topological space $X$, we let $\Delta(X)$ denote the space of Borel probability measures on $X$ endowed with the topology of weak convergence of measures. Every Cartesian product of

topological spaces is endowed with the product topology. A topological space $X$ is metrizable if there is a metric that induces its topology. A Cartesian product of a countable (finite, or denumerable) collection of metrizable spaces is metrizable.

$\Delta_i^1 \subseteq \times_{h_i \in H_i} \Delta\left(Z(h_i)\right)$ is the set of first-order beliefs: the set of $\alpha_i = \left(\alpha_i\left(\cdot|Z(h_i)\right)\right)_{h_i \in H_i}$ such that:

- for all $h_i, h_i' \in H_i$, if $h_i \prec h_i'$ then for every $Y \subseteq Z(h_i')$

$$\alpha_i(Z(h_i')|Z(h_i)) > 0 \Rightarrow \alpha_i\left(Y|Z(h_i')\right) = \frac{\alpha_i(Y|Z(h_i))}{\alpha_i(Z(h_i')|Z(h_i))}; \tag{15}$$

- for all $h \in H$, $a_i \in A_i(h)$, $a_{-i} \in A_{-i}(h)$ (using obvious abbreviations)

$$\alpha_{i,-i}(a_{-i}|h) = \alpha_{i,-i}(a_{-i}|h, a_i). \tag{16}$$

$\Delta_i^2 \subseteq \times_{h_i \in H_i} \Delta\left(Z(h_i) \times \Delta_{-i}^1\right)$ —where $\Delta_{-i}^1 = \times_{j \neq i} \Delta_j^1$— is the set of second-order beliefs, that is, the set of $\beta_i = (\beta_i(\cdot|h_i))_{h_i \in H_i}$ such that:

- if $h_i \prec h_i'$ then

$$\beta_i(h_i'|h_i) > 0 \Rightarrow \beta_i\left(E|h_i'\right) = \frac{\beta_i\left(E|h_i\right)}{\beta_i(h_i'|h_i)} \tag{17}$$

  for all $h_i, h_i' \in H_i$ and every event $E \subseteq Z(h_i') \times \Delta_{-i}^1$;

- $i$'s beliefs satisfy an own-action independence property:

$$\beta_i\left(Z\left(h, (a_i, a_{-i})\right) \times E_\Delta | (h, a_i)\right) = \beta_i\left(Z\left(h, (a_i', a_{-i})\right) \times E_\Delta | (h, a_i')\right), \tag{18}$$

  for every $h \in H$, $a_i, a_i' \in A_i(h)$, $a_{-i} \in A_{-i}(h)$, and (measurable) $E_\Delta \subseteq \Delta_{-i}^1$. The space of second-order beliefs of $i$ is denoted $\Delta_{-i}^2$.

Note that (16) and (18) are given by equalities between marginal measures (on $A_{-i}(h)$ and $A_{-i}(h) \times \Delta_{-i}^1$ respectively).

**Lemma 1** *For each player $i \in I$, $\Delta_i^2$ is a compact metrizable space.*

**Proof** Let $\Theta$ be a non-empty, compact metrizable space. Lemma 1 in Battigalli & Siniscalchi (1999) (B&S) establishes that the set of arrays of probability measures $\left(\mu(\cdot|h_i)\right)_{h_i \in H_i} \in \times_{h_i \in H_i} \Delta\left(Z(h_i) \times \Theta\right)$ such that

$$h_i \prec h_i' \wedge \mu(h_i'|h_i) > 0 \Rightarrow \mu\left(E|h_i'\right) = \frac{\mu\left(E|h_i\right)}{\mu(h_i'|h_i)}$$

is closed. Note that, in the special case where $\Theta$ is a singleton, each $\Delta\left(Z(h_i)\times\Theta\right)$ is isomorphic to $\Delta\left(Z(h_i)\right)$; hence, the set of first-order beliefs satisfying (15) is closed. Letting $\Theta=\Delta^1_{-i}$, we obtain that the set of second-order beliefs satisfying (17) is closed.

Since $\times_{h_i\in H_i}\Delta\left(Z(h_i)\right)$ is a compact subset of a Euclidean space and eq. (16) is a closed condition (equalities between marginal measures are preserved in the limit), Lemma 1 in B&S implies that $\Delta^1_i$ is a closed subset of a compact metrizable space. Hence, $\Delta^1_i$ is a compact metrizable space.

It is well known that if $X_1,...,X_K$ are compact metrizable, so is $\times_{k=1}^K\Delta(X_k)$ (see Aliprantis & Border 2006, Theorem 15.11). Hence, by Lemma 1 in B&S, the set of second-order beliefs satisfying (17) is a closed subset of a compact metrizable space. Since eq. (18) is a closed condition (equalities between marginal measures are preserved in the limit), this implies that $\Delta^2_i$ is compact metrizable. ■

**Lemma 2** *For each profile of behavioral strategies $\sigma=(\sigma_i)_{i\in I}$ there is a unique profile of second-order beliefs $\beta^\sigma=(\beta^\sigma_i)_{i\in I}$ such that $(\sigma,\beta^\sigma)$ is a consistent assessment. The map $\sigma\mapsto\beta^\sigma$ is continuous.*

**Proof** Write $\mathbb{P}^\sigma(h'|h)$ for the probability of reaching $h'$ from $h$, *e.g.*,

$$\mathbb{P}^\sigma(a^1,a^2|\varnothing)=\left(\prod_{j\in I}\sigma_j(a^1_j|\varnothing)\right)\left(\prod_{j\in I}\sigma_j(a^2_j|a^1)\right).$$

Define $\alpha^\sigma_i$ as $\alpha^\sigma_i(z|h)=\mathbb{P}^\sigma(z|h)$ for all $i\in I$, $h\in H$, and $z\in Z$. Define $\beta^\sigma_i$ as $\beta^\sigma_i\left(\cdot|h\right)=\alpha^\sigma_i(\cdot|h)\times\delta_{\alpha^\sigma_{-i}}$ for all $i\in I$, $h\in H$. It can be checked that (1) $\beta^\sigma_i\in\Delta^2_i$ for each $i\in I$, (2) $(\sigma,\beta^\sigma)$ is a consistent assessment, and (3) if $\beta\neq\beta^\sigma$, then either (a) or (b) of the definition of consistency is violated. It is also apparent from the construction that the map $\sigma\mapsto\beta^\sigma$ is continuous, because $\sigma\mapsto\alpha^\sigma$ is obviously continuous, and the Dirac-measure map $\alpha_{-i}\mapsto\delta_{\alpha_{-i}}$ is continuous. ■

**Lemma 3** *The set of consistent assessments is compact.*

**Proof** Lemma 1 implies that $\times_{i\in I}(\Sigma_i\times\Delta^2_i)$ is a compact metrizable space that contains the set of consistent assessments. Therefore, it is enough to show that the latter is closed. Let $(\sigma^n,\beta^n)_{n\in\mathbb{N}}$ be a converging sequence of consistent assessments with limit $(\sigma^\infty,\beta^\infty)$. For each $i\in I$, let $\alpha^n_i$ be the first-order belief derived from $\beta^n_i$ ($n\in\mathbb{N}\cup\{\infty\}$), that is,

$$\alpha^n_i(Y|h)=\beta^n_i(Y\times\Delta^1_{-i}|h)$$

for all $h\in H$ and $Y\subseteq Z(h)$. By consistency, for all $n\in\mathbb{N}$, $i\in I$, $h\in H$, $a\in A(h)$, it holds that

- (a.$n$) $\alpha_i^n(a|h) = \beta_i^n\left(Z(h,a) \times \Delta_{-i}^1|h\right) = \prod_{j\in I} \sigma_j^n(a_j|h)$,
- (b.$n$) $\text{marg}_{\Delta_{-i}^1}\beta_i^n(\cdot|h) = \delta_{\alpha_{-i}^n}$, where each $\alpha_j^n$ is determined as in (a.$n$).

Then,

$$\alpha_i^\infty(a|h) = \beta_i^\infty\left(Z(h,a) \times \Delta_{-i}^1|h\right) = \prod_{j\in I} \sigma_j^\infty(a_j|h)$$

for all $i \in I$, $h \in H$, $a \in A(h)$. Furthermore, $\text{marg}_{\Delta_{-i}^1}\beta_i^\infty(\cdot|h) = \delta_{\alpha_{-i}^\infty}$ for all $i \in I$ and $h \in H$, because $\alpha_{-i}^n \to \alpha_{-i}^\infty$ and the marginalization and Dirac maps $\beta_i \mapsto \text{marg}_{\Delta_{-i}^1}\beta_i$ and $\alpha_{-i} \mapsto \delta_{\alpha_{-i}}$ are continuous. ■

## A.2 Proof of Proposition 1

Fix $i \in I$ arbitrarily. First-order belief $\alpha_i$ is derived from $\beta_i$ and, by consistency, gives the behavioral strategy profile $\sigma$. Therefore, by assumption each $h' \preceq h$ has probability one under $\alpha_i$, which implies that $\mathbb{E}[\pi_i|h';\alpha_i] = \mathbb{E}[\pi_i;\alpha_i]$, hence $\text{F}_i(h';\alpha_i) = 0$. Since blame is capped by frustration, $u_i(h',a_i';\beta_i) = \mathbb{E}[\pi_i|h';\alpha_i]$. Therefore, sequential rationality of the equilibrium assessment implies that $\text{Supp}\sigma_i(\cdot|h') \subseteq \arg\max_{a_i'\in A_i(h')} \mathbb{E}[\pi_i|\,(h',a_i')\,;\alpha_i]$. If there is randomization only in the last stage (or none at all), then players maximize locally their expected material payoff on the equilibrium path. Hence, the second claim follows by inspection of the definitions of agent form of the material-payoff game and Nash equilibrium. ■

## A.3 Proof of Proposition 2

Let $(\bar\sigma,\bar\beta) = \left(\bar\sigma_i,\bar\beta_i\right)_{i\in I}$ be the SE of the material-payoff game, which must be in pure strategies (Remark 3). Fix decision utility functions $u_i(h,a_i;\cdot)$ of the ABI, or ABB kind, and a sequence of real numbers $(\varepsilon_n)_{n\in\mathbb{N}}$, with $\varepsilon_n \to 0$ and $0 < \varepsilon_n < \frac{1}{\max_{i\in I, h\in H}|A_i(h)|}$ for all $n \in \mathbb{N}$. Consider the constrained psychological game where players can choose mixed actions in the following sets:

$$\Sigma_i^n(h) = \{\sigma_i(\cdot|h) \in \Delta(A_i(h)) : \|\sigma_i(\cdot|h) - \bar\sigma_i(\cdot|h)\| \le \varepsilon_n\}$$

if $h$ is on the $\bar\sigma$-path, and

$$\Sigma_i^n(h) = \{\sigma_i(\cdot|h) \in \Delta(A_i(h)) : \forall a_i \in A_i(h), \sigma_i(a_i|h) \ge \varepsilon_n\}$$

if $h$ is off the $\bar\sigma$-path. By construction, these sets are non-empty, convex, and compact. Since the decision utility functions are continuous, and the consistent assessment map $\sigma \longmapsto \beta^\sigma$ is continuous (Lemma 2), the correspondence

$$\sigma \longmapsto \times_{h\in H} \times_{i\in I} \arg\max_{\sigma_i'(\cdot|h)\in\Sigma_i^n(h)} \sum_{a_i\in A_i(h)} \sigma_i'(a_i|h) u_i(h,a_i;\beta_i^\sigma)$$

is upper-hemicontinuous, non-empty, convex, and compact valued; therefore (by Kakutani's theorem), it has a fixed point $\sigma^n$. By Lemma 3, the sequence of consistent assessments $\left(\sigma^n, \beta^{\sigma^n}\right)_{n=1}^{\infty}$ has a limit point $(\sigma^*, \beta^*)$, which is consistent too. By construction, $\bar{\sigma}(\cdot|h) = \sigma^*(\cdot|h)$ for $h$ on the $\bar{\sigma}$-path, therefore $(\bar{\sigma}, \bar{\beta})$ and $(\sigma^*, \beta^*)$ are realization-equivalent. We let $\bar{\alpha}_i$ (respectively, $\alpha_i^*$) denote the first-order beliefs of $i$ implied by $(\bar{\sigma}, \bar{\beta})$ (respectively, $(\sigma^*, \beta^*)$).

We claim that the consistent assessment $(\sigma^*, \beta^*)$ is an SE of the psychological game with decision utility functions $u_i(h, a_i; \cdot)$. We must show that $(\sigma^*, \beta^*)$ satisfies sequential rationality. If $h$ is off the $\bar{\sigma}$-path, sequential rationality is satisfied by construction. Since $\bar{\sigma}$ is deterministic and there are no chance moves, if $h$ is on the $\bar{\sigma}$-path (*i.e.*, on the $\sigma^*$-path) it must have unconditional probability 1 according to each player's beliefs and there cannot be any frustration; hence, $u_i(h, a_i; \beta_i^*) = \mathbb{E}[\pi_i|h, a_i; \alpha_i^*]$ $(i \in I)$ where $\alpha_i^*$ is determined by $\sigma^*$. If, furthermore, it is the second stage ($h = \bar{a}^1$, with $\bar{\sigma}(\bar{a}^1|\varnothing) = 1$), then —by construction— $\mathbb{E}[\pi_i|h, a_i; \alpha_i^*] = \mathbb{E}[\pi_i|h, a_i; \bar{\alpha}_i]$, where $\bar{\alpha}_i$ is determined by $\bar{\sigma}$. Since $\bar{\sigma}$ is an SE of the material-payoff game, sequential rationality is satisfied at $h$. Finally, we claim that $(\sigma^*, \beta^*)$ satisfies sequential rationality also at the root $h = \varnothing$. Let $\iota(h)$ denote the active player at $h$. Since $\iota(\varnothing)$ cannot be frustrated at $\varnothing$, we must show that action $\bar{a}^1$ with $\bar{\sigma}(\bar{a}^1|\varnothing) = 1$ maximizes his expected material payoff given belief $\alpha_{\iota(\varnothing)}$. According to ABB and ABI, player $\iota(a^1)$ can only blame the first mover $\iota(\varnothing)$ and possibly hurt him, if he is frustrated. Therefore, in assessment $(\sigma^*, \beta^*)$ at node $a^1$, either $\iota(a^1)$ plans to choose his (unique) payoff maximizing action, or he blames $\iota(\varnothing)$ strongly enough to give up some material payoff in order to bring down the payoff of $\iota(\varnothing)$. Hence, $\mathbb{E}[\pi_{\iota(\varnothing)}|a^1; \alpha^*_{\iota(a^1)}] \leq \mathbb{E}[\pi_{\iota(\varnothing)}|a^1; \bar{\alpha}_{\iota(a^1)}]$ (anger). By consistency of $(\sigma^*, \beta^*)$ and $(\bar{\sigma}, \bar{\beta})$, $\alpha^*_{\iota(a^1)} = \alpha^*_{\iota(\varnothing)}$ and $\bar{\alpha}_{\iota(a^1)} = \bar{\alpha}_{\iota(\varnothing)}$ (cons.). Since $(\sigma^*, \beta^*)$ is realization-equivalent (r.e.) to $(\bar{\sigma}, \bar{\beta})$, which is the material-payoff equilibrium (m.eq.), for each $a^1 \in A(\varnothing)$,

$$\begin{aligned}
&\mathbb{E}[\pi_{\iota(\varnothing)}|\bar{a}^1; \alpha^*_{\iota(\varnothing)}] \overset{\text{(r.e.)}}{=} \mathbb{E}[\pi_{\iota(\varnothing)}|\bar{a}^1; \bar{\alpha}_{\iota(\varnothing)}] \overset{\text{(m.eq.)}}{\geq} \\
&\mathbb{E}[\pi_{\iota(\varnothing)}|a^1; \bar{\alpha}_{\iota(\varnothing)}] \overset{\text{(cons.)}}{=} \mathbb{E}[\pi_{\iota(\varnothing)}|a^1; \bar{\alpha}_{\iota(a^1)}] \overset{\text{(anger)}}{\geq} \\
&\mathbb{E}[\pi_{\iota(\varnothing)}|a^1; \alpha^*_{\iota(a^1)}] \overset{\text{(cons.)}}{=} \mathbb{E}[\pi_{\iota(\varnothing)}|a^1; \alpha^*_{\iota(\varnothing)}].
\end{aligned}$$

This completes the proof for the ABB and ABI cases. If there are only two players, then we have a leader-follower game and SA is equivalent to ABB, so $(\sigma^*, \beta^*)$ is an SE in this case too. ■

## A.4 Proof of Proposition 3

Recall that $\ell$ and $f$ respectively denote the leader and the follower and that, by convention, the leader has no terminating action. By Remark 3, we obtain the unique and pure material-payoff

SE strategy pair, viz. $(\bar{\sigma}_\ell, \bar{\sigma}_f)$, by backward induction: for each $a_\ell \in A_\ell(\varnothing)$, let

$$\bar{s}_f(a_\ell) = \arg \max_{a_f \in A_f(a_\ell)} \pi_f(a_\ell, a_f)$$

denote the material best reply of $f$ to $a_\ell$ (unique by assumption); then

$$\forall a_\ell \in A_\ell(\varnothing), \bar{\sigma}_f(\bar{s}_f(a_\ell)|a_\ell) = 1,$$

$$\bar{\sigma}_\ell\left(\arg \max_{a_\ell \in A_\ell(\varnothing)} \pi_\ell(a_\ell, \bar{s}_f(a_\ell))|\varnothing\right) = 1. \tag{19}$$

Let $(\bar{a}_\ell, \bar{s}_f)$ denote this pure-strategy equilibrium. We must prove that $\mathbb{E}[\pi_f; \beta] \geq \pi_f(\bar{a}_\ell, \bar{s}_f(\bar{a}_\ell))$ for every SE assessment $(\sigma, \beta)$.

Now, fix arbitrarily an SE $(\sigma, \beta)$, thus, by consistency, $\sigma$ is derived from the first-order beliefs implied by $\beta$. Observe that

$$\forall a_\ell \in A_\ell(\varnothing) \setminus \{\bar{a}_\ell\},\ \mathbb{E}[\pi_\ell|a_\ell; \beta] \overset{\text{(anger)}}{\leq} \pi_\ell(a_\ell, \bar{s}_f(a_\ell)) \overset{\text{(m.eq.)}}{<} \pi_\ell(\bar{a}_\ell, \bar{s}_f(\bar{a}_\ell)). \tag{20}$$

The first inequality holds because either $f$—upon observing $a_\ell$—is angry enough to deviate from his material-payoff maximizing action $\bar{s}_f(a_\ell)$ and punish $\ell$, or he replies with $\bar{s}_f(a_\ell)$; the second inequality holds because $(\bar{a}_\ell, \bar{s}_f)$ is the unique material-payoff equilibrium. Since in every equilibrium the leader, who cannot be frustrated, maximizes his expected material payoff, it must be the case that either (i) $\sigma_\ell(\bar{a}_\ell|\varnothing) = 1$, or (ii) $\mathbb{E}[\pi_\ell|\bar{a}_\ell; \beta] < \pi_\ell(\bar{a}_\ell, \bar{s}_f(\bar{a}_\ell))$, otherwise eq. (20) implies that he would choose $\bar{a}_\ell$ with probability 1. In case (i), $f$ cannot be frustrated after $\bar{a}_\ell$; hence, $\sigma_f(\bar{s}_f(\bar{a}_\ell)|\bar{a}_\ell) = 1$, $\sigma_\ell(\bar{a}_\ell|\varnothing) = 1$, and the equilibrium payoff of $f$ is $\pi_f(\bar{a}_\ell, \bar{s}_f(\bar{a}_\ell))$, the same as in the material-payoff equilibrium. In case (ii), $\sigma_f(\bar{s}_f(\bar{a}_\ell)|\bar{a}_\ell) < 1$, that is, $f$ is not choosing his material-payoff maximizing action $\bar{s}_f(\bar{a}_\ell)$ because he is frustrated. Therefore,

$$\mathbb{E}[\pi_f; \beta] - \pi_f(\bar{a}_\ell, \bar{s}_f(\bar{a}_\ell)) = \mathbb{E}[\pi_f; \beta] - \max_{a_f \in A_f(\bar{a}_\ell)} \pi_f(\bar{a}_\ell, a_f) = \mathrm{F}_f(\bar{a}_\ell; \alpha) > 0,$$

where $\alpha$ is derived from $\beta$. Thus, in each case $\mathbb{E}[\pi_f; \beta] \geq \pi_f(\bar{a}_\ell, \bar{s}_f(\bar{a}_\ell))$. ■

## A.5 Proof of Proposition 4

Recall that $(\bar{a}_\ell, \bar{s}_f)$ denotes the (necessarily pure) material-payoff SE strategy pair of a leader-follower game form with no relevant ties. Let $\hat{s}_f$ be a threat as per Definition 4 and let $\hat{a}_\ell$ denote the (necessarily unique) best response of the leader. We first construct a consistent

assessment $\left(\hat{\sigma},\hat{\beta}\right)$ such that the leader plays $\hat{a}_\ell$ and the follower responds to $\hat{a}_\ell$ with $\hat{s}_f(\hat{a}_\ell)$, then we show that $\left(\hat{\sigma},\hat{\beta}\right)$ is sequentially rational.

Let $\hat{\sigma}$ be such that $\hat{\sigma}_\ell(\hat{a}_\ell)=1$, $\hat{\sigma}_f(\hat{s}_f(\hat{a}_\ell)|\hat{a}_\ell)=1$, let $\hat{\beta}=\beta^{\hat{\sigma}}$ denote the second-order beliefs profile consistent with $\hat{\sigma}$ (see Lemma 2); similarly, we let $\hat{\alpha}$ denote the first-order beliefs consistent with such $\hat{\sigma}$ (and $\hat{\beta}$). The first part of the proof only relies on the on-path features of $\hat{\sigma}$. Then we complete the construction. By Condition 2 of Definition 4,

$$\mathbb{E}[\pi_f;\hat{\alpha}_f]=\pi_f(\hat{a}_\ell,\hat{s}_f(\hat{a}_\ell))=\max_{a_f\in A_f(\hat{a}_\ell)}\pi_f(\hat{a}_\ell,a_f)>\max_{a_f\in A_f(\bar{a}_\ell)}\pi_f(\bar{a}_\ell,a_f)=\pi_f(\bar{a}_\ell,\bar{s}_f(\bar{a}_\ell)).$$

Therefore, the follower's threat $\hat{s}_f$ materially best responds to $\hat{a}_\ell$, and his frustration after the material-payoff SE action $\bar{a}_\ell$ is strictly positive under such beliefs:

$$\mathrm{F}_f(\bar{a}_\ell;\hat{\alpha}_f)=\mathbb{E}[\pi_f;\hat{\alpha}_f]-\max_{a_f\in A_f(\bar{a}_\ell)}\pi_f(\bar{a}_\ell,a_f)=\pi_f(\hat{a}_\ell,\hat{s}_f(\hat{a}_\ell))-\pi_f(\bar{a}_\ell,\bar{s}_f(\bar{a}_\ell))>0.$$

By Remark 4, frustration and blame coincide under ABB, as they do by definition for SA: $\mathrm{F}_f=\mathrm{B}_f$, because frustration is fully blamed on the leader. Therefore, there are a sufficiently high parameter value $\theta_f>0$ and an action $\widetilde{a}_f$ such that

$$\widetilde{a}_f\in\arg\max_{a_f\in A_f(\bar{a}_\ell)}\{\pi_f(\bar{a}_\ell,a_f)-\theta_f\mathrm{B}_f(\bar{a}_\ell;\hat{\alpha}_f)\pi_\ell(\bar{a}_\ell,a_f)\}$$

and

$$\pi_\ell(\bar{a}_\ell,\widetilde{a}_f)\leq\pi_\ell(\bar{a}_\ell,\hat{s}_f(\bar{a}_\ell))\overset{(C1)}{<}\pi_\ell(\hat{a}_\ell,\hat{s}_f(\hat{a}_\ell)),$$

where the latter inequality follows from Condition 1 of Definition 4 (note: $\widetilde{a}_f$ may differ from $\hat{s}_f(\bar{a}_\ell)$ because a frustrated follower may want to punish the leader more than $\hat{s}_f$ does).

With this, we complete the construction of $\hat{\sigma}_f$ (hence of $\hat{\beta}=\beta^{\hat{\sigma}}$ and $\hat{\alpha}$) on top of $\hat{\sigma}_f(\hat{s}_f(\hat{a}_\ell)|\hat{a}_\ell)=1$ as follows:

$$\hat{\sigma}_f(\widetilde{a}_f|\bar{a}_\ell)=1,$$

and

$$\mathrm{supp}\hat{\sigma}_f(\cdot|a_\ell)\subseteq\arg\max_{a_f\in A_f(a_\ell)}\{\pi_f(a_\ell,a_f)-\theta_f\mathrm{B}_f(a_\ell;\hat{\alpha}_f)\pi_\ell(a_\ell,a_f)\}$$

for all the other actions $a_\ell\in A_\ell(\varnothing)\setminus\{\hat{a}_\ell,\bar{a}_\ell\}$. Since the follower with such beliefs is not frustrated after the candidate equilibrium action $\hat{a}_\ell$ of the leader, the construction implies that $\hat{\sigma}_f$ satisfies sequential rationality.

We conclude proving that $\hat{a}_\ell$ is a (material) best response for the leader. For every $a_\ell \in A_\ell(\varnothing)$, we have

$$\begin{aligned}
&\mathbb{E}\left[\pi_\ell | \hat{a}_\ell; \hat{\alpha}_\ell\right] \overset{\text{(constr.)}}{=} \pi_\ell\left(\hat{a}_\ell, \hat{s}_f\left(\hat{a}_\ell\right)\right) \overset{\text{(C2)}}{=} \\
&\pi_\ell\left(\hat{a}_\ell, \bar{s}_f\left(\hat{a}_\ell\right)\right) \overset{\text{(C2)}}{>} \pi_\ell\left(\bar{a}_\ell, \bar{s}_f\left(\bar{a}_\ell\right)\right) \overset{\text{(m.eq)}}{\geq} \\
&\pi_\ell\left(a_\ell, \bar{s}_f\left(a_\ell\right)\right) \overset{\text{(anger)}}{\geq} \mathbb{E}\left[\pi_\ell | a_\ell; \hat{\alpha}_\ell\right],
\end{aligned}$$

where the first equality holds by construction, the second equality and the first inequality follow from Condition 2, and the last two inequalities follow from the fact that $(\bar{a}_\ell, \bar{s}_f)$ is the material-payoff SE, and that the reply the follower under $\hat{\sigma}$ must give (weakly) less to the leader than the material-payoff SE, as an angry follower may trade off some of his material payoff (maximized at $a_f = \bar{s}_f(a_\ell)$) to make the leader's payoff lower than $\pi_\ell\left(a_\ell, \bar{s}_f\left(a_\ell\right)\right)$. ■

## A.6 Proof of Proposition 5

Let $\hat{\sigma} = (\hat{a}_\ell, \hat{s}_f)$ be a pure SE strategy pair with SA/ABB or ABI of a leader-follower game with no relevant ties, and let $\hat{\alpha}$ denote the corresponding profile of first-order beliefs. First note that the consistency condition of SE implies that deviations from $\hat{a}_\ell$ would be rated by $f$ as unintended mistakes, hence $f$ would not be angry. Therefore, there the pure SE with ABI is unique and it coincides with $\hat{\sigma}$. Next we consider SA/ABB and we prove the result by contraposition, that is, we show that if path $\left(\hat{a}_\ell, \hat{s}_f\left(\hat{a}_\ell\right)\right)$ differs from the material-payoff equilibrium path $\left(\bar{a}_\ell, \bar{s}_f\left(\bar{a}_\ell\right)\right)$, then there is a threat.

As the leader cannot be frustrated at the root, and the follower cannot be frustrated after $\hat{a}_\ell$, which is expected with probability 1, we must have $\hat{a}_\ell = r_\ell\left(\hat{s}_f\right)$ and $\hat{s}_f\left(\hat{a}_\ell\right) = \arg\max_{a_f \in A(\hat{a}_\ell)} \pi_f\left(\hat{a}_\ell, a_f\right) = \bar{s}_f\left(\hat{a}_\ell\right)$. Suppose $\left(\hat{a}_\ell, \hat{s}_f\left(\hat{a}_\ell\right)\right) \neq \left(\bar{a}_\ell, \bar{s}_f\left(\bar{a}_\ell\right)\right)$. Then $\hat{a}_\ell \neq \bar{a}_\ell$ and

$$\pi_\ell\left(\hat{a}_\ell, \bar{s}_f\left(\hat{a}_\ell\right)\right) = \pi_\ell\left(\hat{a}_\ell, \hat{s}_f\left(\hat{a}_\ell\right)\right) > \pi_\ell\left(\bar{a}_\ell, \hat{s}_f\left(\bar{a}_\ell\right)\right), \tag{21}$$

because $\hat{a}_\ell = r_\ell\left(\hat{s}_f\right)$. On the other hand, since $(\bar{a}_\ell, \bar{s}_f)$ is the unique material-payoff equilibrium

$$\pi_\ell\left(\hat{a}_\ell, \bar{s}_f\left(\hat{a}_\ell\right)\right) \overset{\text{(m-eq)}}{<} \pi_\ell\left(\bar{a}_\ell, \bar{s}_f\left(\bar{a}_\ell\right)\right).$$

Therefore,

$$\hat{s}_f\left(\bar{a}_\ell\right) \neq \bar{s}_f\left(\bar{a}_\ell\right) = \arg\max_{a_f \in A(\bar{a}_\ell)} \pi_f\left(\bar{a}_\ell, a_f\right),$$

that is, the follower is not choosing his material-payoff best response. This can happen in an SE with SA/ABB only if the follower is frustrated after $\bar{a}_f$, that is,

$$\mathrm{F}_f\left(\bar{a}_\ell; \hat{\alpha}_f\right) = \pi_f\left(\hat{a}_\ell, \hat{s}_f\left(\hat{a}_\ell\right)\right) - \max_{a_f \in A_f(\bar{a}_\ell)} \pi_f\left(\bar{a}_\ell, a_f\right) = \pi_f\left(\hat{a}_\ell, \hat{s}_f\left(\hat{a}_\ell\right)\right) - \pi_f\left(\bar{a}_\ell, \bar{s}_f\left(\bar{a}_\ell\right)\right) > 0.$$

Taking into account that $\hat{s}_f(\hat{a}_\ell) = \bar{s}_f(\hat{a}_\ell)$, this shows that Condition 2 of Definition 4 of threat holds. Furthermore, the material-payoff SE condition for the leader and (21) yield

$$\pi_\ell(\bar{a}_\ell, \bar{s}_f(\bar{a}_\ell)) \overset{\text{(m-eq)}}{\geq} \pi_\ell(\hat{a}_\ell, \bar{s}_f(\hat{a}_\ell)) \overset{(21)}{=} \pi_\ell(\hat{a}_\ell, \hat{s}_f(\hat{a}_\ell)) \overset{(21)}{>} \pi_\ell(\bar{a}_\ell, \hat{s}_f(\bar{a}_\ell)).$$

Therefore Condition 1 holds as well. ■

# References


[1] Aina, C., P. Battigalli, and A. Gamba (2018): "Frustration and Anger in the Ultimatum Game: An Experiment," IGIER working paper 621.

[2] Akerlof, R. J. (2016): "Anger and Enforcement," *Journal of Economic Behavior & Organization,* 126, 110-124.

[3] Alicke, M. D. (2000): "Culpable Control and the Psychology of Blame," *Psychological Bulletin,* 126, 556.

[4] Aliprantis, C. R. and K. C. Border (2006): *Infinite Dimensional Analysis.* Berlin: Springer-Verlag.

[5] Anderson, E. and D. Simester (2010): "Price Stickiness and Customer Antagonism," *Quarterly Journal of Economics,* 125, 729–765.

[6] Arnold, M. B. (1960): *Emotions and Personality.* New York: Columbia University Press.

[7] Attanasi, G., P. Battigalli, and E. Manzoni (2016): "Incomplete Information Models of Guilt Aversion in the Trust Game," *Management Science*, 62, 648 - 667.

[8] Averill, J. R. (1982): *Anger and Aggression: An Essay on Emotion* New York: Springer-Verlag.

[9] Bartling, B. and U. Fischbacher (2012): "Shifting the Blame: On Delegation and Responsibility," *Review of Economic Studies*, 79, 67–87.

[10] Battigalli, P. (1997): "On Rationalizability in Extensive Games," *Journal of Economic Theory*, 74, 40-61.

[11] Battigalli, P., G. Charness, and M. Dufwenberg (2013): "Deception: The Role of Guilt," *Journal of Economic Behavior & Organization*, 93, 227–232.

[12] Battigalli, P., A. Di Tillio, and D. Samet (2013): "Strategies and Interactive Beliefs in Dynamic Games," in D. Acemoglu, M. Arellano, E. Dekel (eds.) *Advances in Economics and Econometrics: Theory and Applications*, Tenth World Congress, Vol. 1 Economic Theory (Econometric Society Monographs No. 49), Cambridge: Cambridge University Press, 391–422.

[13] Battigalli, P. and M. Dufwenberg (2007): "Guilt in Games," *American Economic Review, Papers and Proceedings,* 97, 170–176.

[14] Battigalli, P. and M. Dufwenberg (2009): "Dynamic Psychological Games," *Journal of Economic Theory,* 144, 1–35.

[15] Battigalli, P. and M. Siniscalchi (1999): "Hierarchies of Conditional Beliefs and Interactive Epistemology in Dynamic Games," *Journal of Economic Theory,* 88, 188–230.

[16] Baumeister, R. F. and B. F. Bushman (2007): "Angry Emotions and Aggressive Behaviors," G. Steffgen and M. Gollwitzer (Eds.) *Emotions and Aggressive Behavior*, Cambridge, MA: Hogrefe, 61–75.

[17] Baumeister, R. F., L. Smart, and J. M. Boden (1996): "Relation of Threatened Egotism to Violence and Aggression: The Dark Side of High Self-Esteem," *Psychological Review,* 103, 5–33.

[18] Bell, D. (1985): "Disappointment in Decision Making under Uncertainty," *Operations Research*, 33, 1–27.

[19] Berg, J., J. Dickhaut, and K. McCabe (1995): "Trust, Reciprocity, and Social History," *Games and Economic Behavior* 10, 122–142.

[20] Berkowitz, L. (1978): "Whatever Happened to the Frustration-Aggression Hypothesis?," *American Behavioral Scientist*, 21, 691–708.

[21] Berkowitz, L. (1989): "Frustration-Aggression Hypothesis: Examination and Reformulation," *Psychological Bulletin*, 106, 59–73.

[22] Bolton, G. E., and A. Ockenfels. (1989): "ERC: A Theory of Equity, Reciprocity, and Competition," *American Economic Review*, 90, 166–193.

[23] Bosman, R., M. Sutter, and F. van Winden (2005): "The Impact of Real Effort and Emotions in the Power-to-take Game," *Journal of Economic Psychology,* 26, 407–429.

[24] Bosman, R. and F. van Winden (2002): "Emotional Hazard in a Power-to-take Experiment," *Economic Journal*, 112, 147–169.

[25] Brandts, J., and C. Sola (2001): "Reference Points and Negative Reciprocity in Simple Sequential Games," *Games and Economic Behavior*, 36, 138-157.

[26] Buss, D. (2016): *Evolutionary Psychology. The New Science of the Mind*, New York: Routledge (5th edition).

[27] Brams, S. (2011): *Game Theory and the Humanities: Bridging Two Worlds*, Cambridge, MA: MIT Press.

[28] Card, D. and G. Dahl (2011): "Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior," *Quarterly Journal of Economics,* 126, 103–143.

[29] Carpenter, J. and P. Matthews (2012): "Norm Enforcement: Anger, Indignation or Reciprocity," *Journal of the European Economic Association,* 10, 555–572.

[30] Celen, B., A. Schotter, and M. Blanco (2017): "On Blame and Reciprocity: An Experimental Study," *Journal of Economic Theory,* 169, 62–92.

[31] Chang, L., A. Smith, M. Dufwenberg, and A. Sanfey (2011): "Triangulating the Neural, Psychological, and Economic Bases of Guilt Aversion," *Neuron,* 70, 560–572.

[32] Charness, G., and M. Rabin (2002): "Understanding Social Preferences With Simple Tests, " *The Quarterly Journal of Economics*, 117, 817-869.

[33] Dollard, J., L. Doob, N. Miller, O. Mowrer, and R. Sears (1939): *Frustration and Aggression.* New Haven, NJ: Yale University Press.

[34] Dufwenberg, M. (2002): "Marital Investment, Time Consistency and Emotions," *Journal of Economic Behavior and Organization* 48, 57–69.

[35] Dufwenberg, M. and G. Kirchsteiger (2004): "A Theory of Sequential Reciprocity," *Games and Economic Behavior*, 47, 268–298.

[36] Dufwenberg, M., F. Li, and A. Smith (2018*a*): "Promises and Punishment," unpublished manuscript.

[37] Dufwenberg, M., F. Li, and A. Smith (2018*b*): "Threats," unpublished manuscript.

[38] Dufwenberg, M., A. Smith, and M. Van Essen (2013): "Hold-up: With a Vengeance," *Economic Inquiry*, 51, 896–908.

[39] Elster, J. (1998): "Emotions and Economic Theory, " *Journal of Economic Literature*, 36, 47-74.

[40] Falk, A. and U. Fischbacher (2006): "A Theory of Reciprocity," *Games and Economic Behavior*, 54, 293–315.

[41] Falk, A., E. Fehr, and U. Fischbacher (2003): "On the Nature of Fair Behavior," *Economic Inquiry,* 41, 20-26.

[42] Falk, A., E. Fehr, and U. Fischbacher (2008): "Testing Theories of Fairness—Intentions Matter," *Games and Economic Behavior,* 62(1), 287-303.

[43] Fehr, E. and S. Gaechter (2002): "Altruistic Punishment in Humans," *Nature*, 415, 137–140.

[44] Fehr, E. and K. M. Schmidt (1999): "A Theory of Fairness, Competition, and Cooperation," *The Quarterly Journal of Economics*, 114, 817-868.

[45] Frank, R. H. (1988): *Passions Within Reason: The Strategic Role of the Emotions.* Chicago: W.W. Norton & Co.

[46] Frijda, N. H. (1986): *The Emotions.* Cambridge: Cambridge University Press.

[47] Frijda, N. H. (1993): "The Place of Appraisal in Emotion," *Cognition and Emotion*, 7, 357–387.

[48] Gale, J., K. G. Binmore, and L. Samuelson (1995): "Learning to be Imperfect: The Ultimatum Game," *Games and Economic Behavior*, 8, 56–90.

[49] Geanakoplos, J., D. Pearce, and E. Stacchetti (1989): "Psychological Games and Sequential Rationality," *Games and Economic Behavior*, 1, 60–79.

[50] Gneezy, U. and A. Imas (2014): "Materazzi Effect and the Strategic Use of Anger in Competitive Interactions," *Proceedings of the National Academy of Sciences,* 111, 1334–1337.

[51] Grimm, V. and F. Mengel (2011): "Let Me Sleep on It: Delay Reduces Rejection Rates in Ultimatum Games," *Economics Letters*, 111, 113–115.

[52] Güth, W. and H. Kliemt (1998): "The Indirect Evolutionary Approach: Bridging between Rationality and Adaptation," *Rationality and Society*, 10, 377–399.

[53] Guth, W., R. Schmittberger, and B. Schwarze (1982): "An Experimental-Analysis of Ultimatum Bargaining," *Journal of Economic Behavior and Organiztion* 3, 367–388.

[54] Gurdal, M., J. Miller, and A. Rustichini (2014): "Why Blame?" *Journal of Political Economy,* 121, 1205–1246.

[55] Harmon-Jones, E. and J. Sigelman (2001): "State Anger and Prefrontal Brain Activity: Evidence that Insult-Related Relative Left-Prefrontal Activation is Associated with Experienced Anger and Aggression," *Journal of Personality and Social Psychology*, 80, 797–803.

[56] Halpern, J. Y. (2016): *Actual Causality.* Cambridge: MIT Press.

[57] Hirshleifer, J. (1987): "On the Emotions as Guarantors of Threats and Promises," John Dupre (Ed.), *The Latest on the Best: Essays on Evolution and Optimality*, Cambridge: MIT Press, Chapter 14, 307–26.

[58] Klein, D. B. and B. O,Flaherty (1993): "A Game-Theoretic Rendering of Promises and Threats," *Journal of Economic Behavior & Organization,* 21, 295–314.

[59] Kőszegi, B. and M. Rabin (2006): "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics,* 121, 1133–1166.

[60] Kőszegi, B. and M. Rabin (2007): "Reference-Dependent Risk Attitudes," *American Economic Review,* 97, 1047–1073.

[61] Kreps, D. and R. Wilson (1982): "Sequential Equilibria," *Econometrica*, 50, 863–894.

[62] van Leeuwen, B., C. Noussair, T. Offerman, S. Suetens, M. van Veelen, and J. van de Ven (2018): "Predictably Angry – Facial Cues Provide a Credible Signal of Destructive Behavior," *Management Science*, 64, 2973-3468.

[63] Loomes, G. and R. Sugden (1986): "Disappointment and Dynamic Consistency in Choice under Uncertainty," *Review of Economic Studies*, 53, 271–282.

[64] Marcus-Newhall, A., W.C. Pedersen, M. Carlson, and N. Miller (2000): "Displaced Aggression is Alive and Well: a Meta-Analytic Review," *Journal of Personality and Social Psychology*, 78, 670—689.

[65] Munyo, I., and M.A. Rossi (2013): "Frustration, Euphoria, and Violent Crime," *Journal of Economic Behavior & Organization*, 89, 136-142.

[66] Nelson, W. R. (2002): "Equity or Intention: It Is the Thought That Counts," *Journal of Economic Behavior & Organization*, 48, 423–430.

[67] Oechssler, J., A. Roider, and P.W. Schmitz (2015): "Cooling Off in Negotiations: Does it Work?," *Journal of Institutional and Theoretical Economics*, 171, 565–588.

[68] Passarelli, F. and G. Tabellini (2017): "Emotions and Political Unrest," *Journal of Political Economy*, 125, 903 - 946.

[69] Persson, E. (2018): "Testing the Impact of Frustration and Anger When Responsibility is Low," *Journal of Economic Behavior and Organization,* 145, 435-448.

[70] Pillutla, M. and K. Murningham (1996): "Unfairness, Anger, and Spite: Emotional Rejections of Ultimatum Offers," *Organizational Behavior and Human Decision Processes*, 68, 208–224.

[71] Potegal, M., C. Spielberger, and G. Stemmler (eds.) (2010): *International Handbook of Anger: Constituent and Concomitant Biological, Psychological, and Social Processes*. New York: Springer-Verlag.

[72] Rabin, M. (1993): "Incorporating Fairness into Game Theory and Economics," *American Economic Review*, 83, 1281–1302.

[73] Rauh, M. and G. Seccia (2006): "Anxiety and Performance: A Learning-by-Doing Model," *International Economic Review*, 47, 583–609.

[74] Reuben, E. and F. van Winden (2008): "Social Ties and Coordination on Negative Reciprocity: The Role of Affect," *Journal of Public Economics,* 92, 34–53.

[75] Rotemberg, J. (2005): "Customer Anger at Price Increases, Changes in the Frequency of Price Adjustment and Monetary Policy," *Journal of Monetary Economics,* 52, 829–852.

[76] Rotemberg, J. (2008): "Minimally Acceptable Altruism and the Ultimatum Game," *Journal of Economic Behavior and Organization,* 66, 457–476.

[77] Rotemberg, J. (2011): "Fair Pricing," *Journal of the European Economic Association,* 9, 952–981.

[78] Sanfey, A. (2009): "Expectations and Social Decision-Making: Biasing Effects of Prior Knowledge on Ultimatum Responses," *Mind and Society* 8, 93–107.

[79] Schotter, A. and B. Sopher (2007): "Advice and Behavior in Intergenerational Ultimatum Games: An Experimental Approach," *Games and Economic Behavior*, 58, 365–393.

[80] Sell, A., L. Cosmides, J. Tooby (2009): "Formidability and the Logic of Human Anger," *Proceedings of the National Academy of Sciences*,106, 15073-15078.

[81] Selten, R. (1975): "Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 4, 25–55.

[82] Selten, R. (1978): "The chain store paradox," *Theory and Decision*, 9, 127–159.

[83] Silfver, M. (2007): "Coping with Guilt and Shame: A Narrative Approach," *Journal of Moral Education,* 36, 169–183.

[84] Smith, A. (2009): "Belief-Dependent Anger in Games," typescript, University of Arizona.

[85] Sutter, M. (2007): "Outcomes Versus Intentions: On the Nature of Fair Behavior and Its Development With Age," *Journal of Economic Psychology*, 28, 69–78.

[86] Winter, E. (2014): *Feeling Smart: Why Our Emotions Are More Rational Than We Think.* New York: PublicAffairs.

[87] Winter, E., I. Garcia-Jurado, and L. Mendez Naya (2016): "Mental Equilibrium and Strategic Emotions," *Management Science*, 63(5), 1302–1317.

[88] Xiang, T., T. Lohrenz, and R. Montague (2013): "Computational Substrates of Norms and Their Violations during Social Exchange," *Journal of Neuroscience*, 33, 1099–1108.