Procedural Concerns in Psychological Games*

Alexander Sebald†

September 28, 2007

Abstract

One persistent finding in experimental economics is that people react very differently to outcomewise identical situations depending on the procedures which have led to them. In accordance with this, there exists a broad consensus among psychologists that not only expected outcomes shape human behavior, but also the way in which decisions are taken. Economists, on the other hand, have remained remarkably silent about procedural aspects of strategic interactions. This paper provides a game theoretic framework that integrates procedural concerns into economic analysis. Building on Battigalli and Dufwenberg (2005)'s framework of dynamic psychological games, we show how procedural concerns can be conceptualized assuming that agents are (also) motivated by belief-dependent psychological payoffs. Procedural choices influence the causal attribution of responsibilities, the evaluation of intentions and the arousal of emotions. Two applications highlight the impact and importance of procedural concerns in strategic interactions.

Keywords: Psychological Game Theory, Procedural Concerns, Reciprocity, Guilt Aversion

JEL Classification: D01, C70

1 Introduction

One persistent finding in experimental economics is that people react very differently to outcomewise identical situations depending on the *procedures* which have led to them [e.g. Blount (1995), Falk et al (2000), Charness (2004), Brandts et al (2006), Charness and Levine (2007)]. For example, Charness and Levine (2007) experimentally analyze workers' reactions to pay decisions by firms following different

*I am very grateful to Estelle Cantillon, Paolo Casini, Gary Charness, Paola Conconi, Werner Güth and Georg Kirchsteiger for helpful comments.

†Department of Economics, University of Maastricht, PO Box 616, 6200 MD Maastricht, The Netherlands, and ECARES, Université Libre de Bruxelles. Sebald is also member of ECORE, the recently created association between CORE and ECARES. E-mail: a.sebald@algec.unimaas.nl

wage-setting procedures.[1] They find that the process leading to a specific wage affects the workers' effort choice. Given the same wage, workers choose significantly more often low effort when the wage-setting process reveals less-good intentions by firms compared to situations in which intentions are perceived as good. In the same spirit, Brandts et al (2006) show that selection procedures matter in a three-player game in which one player has to select one of the other players to perform a specific task.[2] In their experiment selected players behave very differently in their subsequent tasks depending on the type of procedure which was used to select them. They suggest that people exhibit *procedural concerns* because selection procedures affect the beliefs that people hold about each others' intentions and expectations which subsequently influence their behaviors.

This paper provides a game theoretic framework that integrates procedural concerns into economic analysis. Building on Battigalli and Dufwenberg (2005)'s *dynamic psychological games*, we show how procedural concerns can be conceptualized assuming that agents are (also) motivated by belief-dependent psychological payoffs. Our paper consist of three building blocks: a class of *procedural games* in which agents choose for procedures rather than for actions as traditionally assumed in game theory, agents with belief-dependent utilities as defined by Battigalli and Dufwenberg (2005) and a solution concept, sequential psychological equilibrium. Using these three building blocks we show how procedural choices influence the causal attribution of responsibilities, the evaluation of intentions and the arousal of emotions.

Among psychologists there exists by now a broad consensus that not only expected outcomes shape human behavior, but also the way in which decisions are taken [e.g. Thibaut and Walker (1975), Lind and Tyler (1988), Collie et al (2002), Anderson and Otto (2003) and Blader and Tyler (2003)]. Prominent examples of areas in which procedures have been found to play an eminent role are workplace relations and the public acceptability of policies and laws. Psychologists have found evidence that behavioral reactions to promotion decisions, bonus allocations, dismissals etc. strongly depend on the perceived fairness of selection procedures [e.g. Lemons and Jones (2001), Konovsky (2000), Bies and Tyler (1993), Lind et al (2000) and Roberts and Markel (2001)] and that public compliance with policies and laws strongly depends on the perceived fairness of their enforcement procedures [e.g. Tyler (1990), Wenzel (2002), Murphy (2004), De Cremer and van Knippenberg (2003) and Tyler (2003)].

[1] In their experiment firms have to choose between either a low ($4) or a high ($8) wage. Following the firm's decision, this wage is either decreased or increased by $2 depending on a stochastically determined (i.e. flip of a coin) economic condition which can either be good or bad. After the revelation of the economic condition, the workers have to choose their effort level: low, medium or high. The flip of the coin introduces the possibility to compare to different intentional states that represent two different ways through which the same wage, i.e. $6, is determined: *i*) good intentions: high wage coupled with bad economic condition and *ii*) less-good intentions: less costly low wage coupled with good economic condition.

[2] Two different treatments are studied which differ with regard to the selection procedure. In both treatments the task of the selected player is to choose between two different payoff allocations determining the payoff of all three players.

Psychologists explain the impact of procedures on human interactions with the help of *attribution theory* [e.g. Heider (1958), Kelley (1967), Kelley (1973), Ross and Fletcher (1985)]. Attribution theory assumes that people need to infer causes and assign responsibilities for why outcomes occur. It is argued that especially when outcomes are unfavorable and perceptions of intentions are strong, there is a tendency to assign responsibility for outcomes to people. The assignment of responsibility and blame in turn has been shown to affect the occurrence and intensity of emotions like disappointment, guilt, anger and aggression [Blount (1995)]. To exemplify, imagine a workplace situation in which a principal wants to promote one out of two agents. If he chooses to take the decision on who is to be promoted intransparently, e.g. behind closed doors, agents are driven to attach a high degree of responsibility for the outcome to the principal. His choice is interpreted as intentional, which fosters perceptions of favoritism. If, by contrast, the principal uses a transparent procedure which credibly shows that the decision is based on an unbiased criterion, i.e. a criterion which a priori ensures that both agents have the same chance to be promoted, the principal is not blamed for the final outcome. Hence, if the agents care about intentions their reaction to the same promotion decision will differ depending on the promotion procedure used by the principal. Hence, according to the psychological literature procedures influence the responsibility that people have for specific outcomes, they mitigate the evaluation of intentions and subsequent behaviors.

Notwithstanding the experimental and psychological evidence and the fact that e.g. workplace relations also play an eminent role in the economic literature, traditional economic theory has remained remarkably silent about the impact of procedures on human behavior. Only three recent economic papers have started to theoretically address the issue of procedural concerns [Bolton et al (2005), Trautmann (2006), Krawczyk (2007)]. In contrast to the psychologists' view, however, they all extend models of distributional preferences to account for the impact of procedural choices on strategic interactions. Bolton et al (2005) and Krawczyk (2007)'s models are based on the theory of inequity aversion by Bolton and Ockenfels (2000). Trautmann, on the other hand, builds on Fehr and Schmidt (1999)'s model of social concerns. All three take a similar approach suggesting that the experimental evidence on procedural concerns can be accounted for when agents' utilities depend on expected outcome differences ex ante as well as ex post to any outcome realization.

As indicated in the beginning, Brandts et al (2006) and Charness and Levine (2007) follow the the psychologists' view. They argue that intention-based models, e.g. models of reciprocity and guilt aversion, rather than distributional preferences, explain the experimental evidence on procedural concerns.

Economic theory has widely neglected emotions and intentions as these issues are difficult to reconcile with the traditional presumption of stable consequentialist preferences. Spurred by experimental findings, economists have only recently started to look at the impact of belief-dependent motivations on strategic interactions. Departing from the strictly consequentialist tradition in economics e.g. Geanakoplos et al (1989) and Battigalli and Dufwenberg (2005) have developed a framework to analyze the strategic interaction of agents with belief-dependent motivations: psychological game theory. Roughly speaking, psychological games are games in which agents are (also) motivated by belief-dependent psychological payoffs capturing their emotional involvement. Emotions depend on beliefs about intentions [Elster (1998)]. Geanakoplos et al (1989) concentrate on games in which only agents' initial beliefs matter, whereas Battigalli and Dufwenberg (2005) develop a dynamic framework in which agents update their beliefs about their own and the others' intentions as games unfold. Many types of emotions (e.g. regret, disappointment, guilt, reciprocity) have already been formalized in the context of psychological games. Rabin (1993), Dufwenberg and Kirchsteiger (2004) and Falk and Fischbacher (2006), for example, analyze the strategic interaction of agents that act reciprocally. Battigalli and Dufwenberg (2006) look at the interaction of agents that feel guilt, i.e. that are guilt averse.

Although all of these models are able to explain observed behaviors in experiments in contradiction to classical assumptions about human conduct [e.g. Charness and Dufwenberg (2006), Charness and Rabin (2002), Fehr and Gaechter (2000)], none of them explores the role of procedural choices in the interaction of *emotional agents*.[3] Therefore, different from the existing literature on psychological games, in this paper we concentrate on the impact of procedural choices on the interaction of agents with belief-dependent motivations. First, we show that procedural concerns can theoretically be conceptualized assuming that agents are (also) incentivized by belief-dependent psychological payoffs. As procedural choices affect the beliefs that people hold and agents utilities are belief-dependent, emotional agents exhibit procedural concerns. Second, we show that the behavioral predictions of the already existing literature on psychological games are sensitive to the availability of different procedures to take the same decision. In the existing literature on psychological games it is implicitly assumed that people can only use procedures that make them fully responsible for the outcomes of their actions. In our procedural games people can choose between different procedures to take the same decision. As will be seen, this leads to different equilibrium predictions compared to the existing literature on psychological games. In another paper [Sebald (2007)] it was already shown how procedural concerns affect the strategic interaction of reciprocal agents. Sebald (2007), thus, is an application of the general framework presented here.

Our work is related to the (experimental) literature on the impact of *institutions* on human interaction [North (1991), Bowles (1998), Bohnet (2006), Bohnet (2007)]. In this literature institutions are commonly defined as humanly devised *rules of the game* that structure political, economic and social interactions. The argument is that institutions create and direct incentives, affect preferences, provide information on processes leading to certain outcomes and allow people to make inferences about others' motivations [Bohnet (2006)]. Bohnet and Zeckhauser (2004) and Hong and Bohnet (2005), for example, experimentally investigate the effect of causal attribu-

[3]Following Elster (1998), throughout the paper we will sometimes refer to agents with belief-dependent psychological payoffs as emotional agents.

tion in different institutional settings. They analyze the first-mover behavior in two closely related but different institutional environments, a binary-choice trust and a binary-choice risky dictator game.[4] Participants act differently in the two settings suggesting that people dislike being betrayed by others more than losing a lottery. This implies that there is an additional psychological factor influencing the strategic interaction related to the causal attribution of responsibilities [Bohnet (2006)]. In line with this, our methodological approach sheds light on the hidden relation connecting the information on procedures entailed in institutions and the process of causal attribution. Our work suggests that the process information entailed in institutions creates the possibility for causal attribution and directs it in such a way that people are only held accountable for what they are actually responsible.

The organization of the paper is as follows: In the next section we formally define procedures and characterize a class of *procedural games* in which agents choose procedures rather than actions and strategies. In section 3 we study the impact of procedures on the behavior of emotional agents. More precisely, we characterize agents with belief-dependent psychological payoffs in the context of our class of procedural games and provide a first example of the impact of procedural choices on their strategic interactions. In section 4 we develop the concept *sequential psychological equilibrium* for our procedural games with psychological incentives. Finally, we discuss two applications that highlight the impact and importance of procedural concerns in strategic interactions of reciprocal and guilt averse agents.

2 Procedures and Procedural Games

In this section we proceed in two steps. First, we intuitively sketch our methodological approach with the help of two examples. In a second step we formally define the concept of procedures and fully characterize our class of procedural games in which agents do not choose actions and strategies, as usually assumed in game theory, but procedures. This class of multi-stage games is used in the subsequent sections to capture and analyze the impact of procedural choices on the strategic interaction of agents with belief-dependent utilities.

As a starting point consider games Γ_1 and Γ_2 in Figure 1 and 2:

[Figure 1 and 2 here]

The sole difference between games Γ_1 and Γ_2 is that in Γ_2 player 1 can choose (M) on top of his pure actions (L) and (R). Player 1's pure action (M), however, is nothing

[4] In both situations the first-mover has to decide between either an outside option or to let a second-mover decide between two alternative payoff allocations. In the binary-choice trust game the second-mover is another player. In the binary-choice risky dictator game, on the other hand, the second-mover is *chance* reducing the role of the second player to being a dummy. It is found that first-movers act differently if the responder is the other player compared to the situation in which *chance* acts as the second-mover.

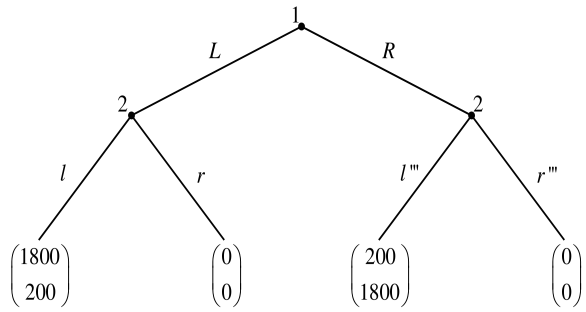

Figure 1: Game Γ_1

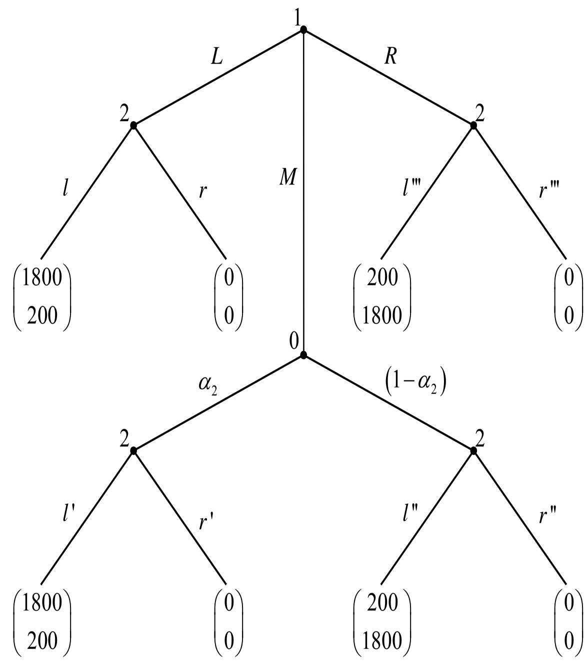

Figure 2: Game Γ_2

else than choosing an explicit randomization device, (0), assigning probabilities α_2 and $(1-\alpha_2)$ to his pure actions (L) and (R) respectively. *Flipping a coin* constitutes such an explicit randomization device, for example. It assigns the probability $\frac{1}{2}$ to both pure actions (L) and (R). Obviously, flipping a coin is just one way in which a decision can be taken. In reality, one usually disposes of many different *credible ways*. Consider, for example, the workplace situation sketched in the introduction. The principal could take the promotion decisions by organizing a promotion tournament using an objective evaluation criterion. Given that both agents are identical, i.e. are equally skilled, and this is commonly known, this would induce a commonly known probability distribution over the set of pure actions giving both agents an equal *chance* to be promoted. Note that we distinguish between explicit, i.e. credible, randomizations which are observed by all players and implicit randomizations, i.e. behavioral strategies. The choice (M) of player 1 in game Γ_2 is a pure choice for an explicit randomization device and it differs from player 1 choosing a behavioral strategy in game Γ_1 which implicitly randomizes over his pure actions (L) and (R) without the others observing the random draw.

But not only choices like (M) can be formalized as choices for explicit randomization devices. Taking the thought about the *credible ways* to the extreme shows that also pure actions like (L) and (R) can equally be defined as choices for explicit randomization mechanisms. Imagine, for example, that player 1 in Γ_1 and Γ_2 chooses his pure action (L). This is equivalent to saying that player 1 chooses for *chance* to take the decision between (L) and (R) assigning probability 1 to his pure action (L). Hence, although (L) represents a pure action, it can nevertheless be reinterpreted in a way in which the decision is indirectly taken by *chance* randomizing with a degenerated probability distribution over the set $\{(L), (R)\}$. This shows that in our two examples, Γ_1 and Γ_2, any choice for a pure action, i.e. (L) and (R), and any choice for an explicit randomization mechanism, i.e.(M), can likewise be reinterpreted as a choice for an explicit randomization device through which the actual decision is subsequently taken by *chance*.

Consider, for example, game Γ_3 in Figure 3, which is a restatement of game Γ_2 in the spirit of this intuition:[5]

[Figure 3 here]

As one can see, in Γ_3 we reformulate all strategic choices of game Γ_2 into choices for explicit randomization mechanisms through which decisions are subsequently taken. In game Γ_2 player 1 can decide between (L), (M) and (R). Equivalently, in game Γ_3 he has to decide between the explicit randomization devices ω_{1,h^0}, ω'_{1,h^0} and ω''_{1,h^0} in the initial history h^0. First, by choosing ω_{1,h^0} he decides to let *chance* take the decision between (L) and (R) assigning probability 1 to (L), i.e. $\rho(L) = 1$. Second, by choosing ω'_{1,h^0} he decides to let *chance* take the decision between (L) and (R) assigning probability α_2 to (L), i.e. $\rho(L) = \alpha_2$, and $(1-\alpha_2)$ to (R), i.e. $\rho(R) = (1-\alpha_2)$.

[5]Note, actions that are played by player *chance* with probability 0 are disregarded in Figure 3.

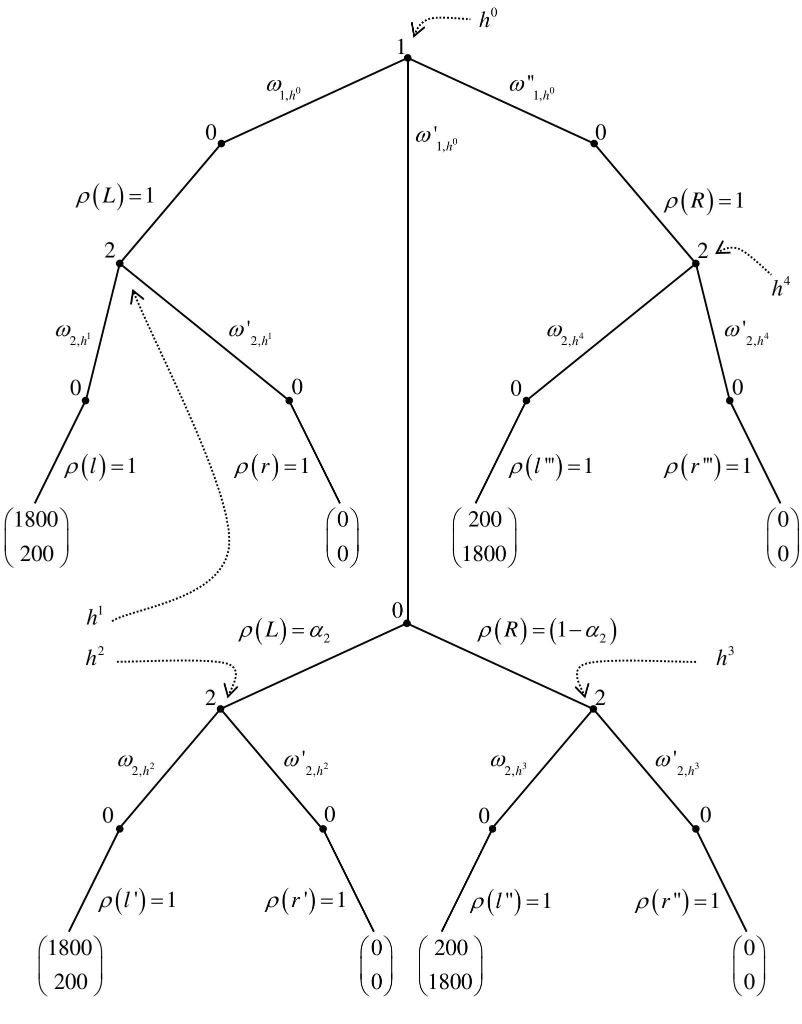

Figure 3: Game Γ_3

Finally, by choosing ω''_{1,h^0} he decides to let *chance* take the decision between (L) and (R) assigning probability 1 to (R), i.e. $\rho(R) = 1$. In all these three cases player 1 only determines how *chance* subsequently takes the decision, rather then taking the decision himself.

Hence, despite the equivalence between games Γ_2 and Γ_3, an interpretive difference exists. Choosing for an explicit randomization mechanism implies that players do not take decisions themselves. They merely determine how decisions are taken by *chance*. In other words, players decide about the procedures which are used to take decisions. The example in Figure 3, thus, uncovers that strategic decision making is not only about choosing actions but also about *how* actions are chosen. For this reason we call game Γ_3 a *procedural game*.

This brings us to the formal definition of our class of procedural games. Formally, let the set of players be $\mathcal{N} = \{0, 1, ..., N\}$ where 0 denotes the uninterested player *chance*. Denote as $\mathcal{H}$ the finite set of histories h, with the empty sequence $h^0 \in \mathcal{H}$, and Z the set of end nodes. Histories $h \in \mathcal{H}$ are sequences that describe the choices that players have made on the path to history h. We assume that only one player moves after each non-terminal history. Hence, the set of histories $\mathcal{H}\backslash\{Z\}$ can be partitioned into sets $\mathcal{H}_i$, with $i \in \mathcal{N}$. At each non-terminal history $h \in \mathcal{H}_i$ after which player $i \in \mathcal{N}\backslash\{0\}$ has to move he disposes of a finite set of feasible actions denoted by $\mathcal{A}_i(h)$ and a finite set of explicit randomization devices denoted by $\Omega_i(h)$ through which he can indirectly choose an action from $\mathcal{A}_i(h)$. In fact, as already suggested in example Γ_3, in our procedural games players $i \in \mathcal{N}\backslash\{0\}$ do not choose actions $a_{i,h} \in \mathcal{A}_i(h)$ directly, but choose explicit randomization mechanisms, denoted by $\omega_{i,h} \in \Omega_i(h)$, through which a decision is indirectly taken by *chance*. The choice for a specific explicit randomization device $\omega_{i,h}$ in history h by player $i \in \mathcal{N}\backslash\{0\}$ determines the explicit probability distribution $\rho_{0,h'}$ with which *chance* takes the actual decision in the following history $h' = (h, \omega_{i,h})$. Hence, any history h controlled by a player $i \in \mathcal{N}\backslash\{0\}$ is succeeded by a history h' controlled by player 0. More formally, if player $i \neq 0$ chooses $\omega_{i,h}$ in history h with length x, then player 0 takes a decision $a_{0,h'}$ in history $h' = (h, \omega_{i,h})$ of length $x + 1$ explicitly randomizing with the probability distribution $\rho_{0,h'}$ over the set $\mathcal{A}_0(h') = \mathcal{A}_i(h)$.[6]

To exemplify, the choice for a pure action (e.g. (L) in Γ_2) translates in our procedural game into a choice for an explicit randomization mechanisms $\omega_{i,h}$ that is associated with a degenerated probability distribution $\rho_{0,h'}$ which assigns probability 1 to the pure action $a_{i,h}$ in the set of possible actions $\mathcal{A}_0(h') = \mathcal{A}_i(h)$. The choice for an explicit randomization device like e.g. (M) in Γ_2), on the other hand, is a choice for an explicit randomization mechanism, $\omega'_{i,h}$, that is associated with a non-degenerate probability distribution $\rho'_{0,h'}$ defined on $\mathcal{A}_0(h') = \mathcal{A}_i(h)$.

This means, player *chance* essentially plays a commonly known, i.e. explicit, mixed strategy $\rho_0 = (\rho_{0,h})_{h\in\mathcal{H}_0}$ which specifies for each history $h \in \mathcal{H}_0$ that he

[6]Note that the length of a history corresponds to the number of choices that are contained in that history.

controls a behavioral strategy $\rho_{0,h}$ according to which an action $a_{0,h}$ is chosen from $\mathcal{A}_0(h)$. Consequently, one can denote as $\rho_0(s_0|h)$ the probability with which player 0 plays the pure strategy $s_0 = (a_{0,h})_{h \in \mathcal{H}_0}$ conditional on history h.

Intuitively, as players only decide on *how* decisions are taken, they only decide on the procedures, which are used to take them. This brings us to a formal definition of procedures:

Definition 1 *A procedure, $\omega_{i,h} \in \Omega_i(h)$, for player $i \in \mathcal{N} \setminus \{0\}$ in history $h \in \mathcal{H}_i$ is a tuple:*[7]

$$\langle \rho_{0,h'}, \mathcal{A}_0(h') \rangle ,$$

where $h' = (h, \omega_{i,h})$ and $\rho_{0,h'}$ is an explicit probability distribution defined on $\mathcal{A}_0(h') = \mathcal{A}_i(h)$.

For a given set of procedures $\Omega_i(h)$, the associated set of explicit probability distributions is denoted by $\mathcal{P}_i(h) = \{\rho_{0,h'} \mid \omega_{i,h} \in \Omega_i(h)\}$. The minimum number of procedures that a player $i \in \mathcal{N} \setminus \{0\}$ can decide between in any history h controlled by him equals the number of pure actions that he has in the traditional extensive form representation.

We define a *procedural strategy* for player $i \in \mathcal{N} \setminus \{0\}$ as a collection that specifies a procedure for each history $h \in \mathcal{H}_i$ after which player i moves, $\omega_i = (\omega_{i,h})_{h \in \mathcal{H}_i}$, where $\omega_{i,h}$ is the procedure that would be selected by player i if h occurred. It is assumed that all players learn the outcome of a procedure directly after its realization and perfect recall holds.

Let $\Omega_i = \times_{h \in \mathcal{H}_i} \Omega_i(h)$ and $\Omega = \times_{\mathcal{N} \setminus \{0\}} \Omega_i$. Given a procedural strategy, $\omega_i \in \Omega_i$ for each player $i \in \mathcal{N} \setminus \{0\}$ and the commonly known system of probability distributions, $\mathcal{P} = \cup_{i \in \mathcal{N} \setminus \{0\}} \mathcal{P}_i$, where $\mathcal{P}_i = \cup_{h \in \mathcal{H}} \mathcal{P}_i(h)$, we can compute a probability distribution over end nodes. By assigning payoffs to end nodes, we can derive an expected payoff function, $\pi_i : Z \times \mathcal{P} \rightarrow \Re$, for every player $i \in \mathcal{N} \setminus \{0\}$ which depends on what *procedural profile*, $\omega \in \Omega$ is played. In what follows, we assume that payoffs are material payoffs like money or any other measurable quantity of some good.

Summarizing:

Definition 2 *A procedural game is a tuple:*

$$\Gamma = \left\langle \mathcal{N}, \Omega, (\pi_i : Z \times \mathcal{P} \rightarrow \Re)_{\mathcal{N} \setminus \{0\}} \right\rangle .$$

This concludes the definition of procedures and the characterization of our class of procedural games which is the basis for our subsequent analysis. Starting from two

[7] In example Γ_3 procedures are used to choose pure actions. We do not exclude, however, the possibility that players use procedures to choose between procedures and procedures that choose between procedures that choose between procedures etc. Procedures, $\omega_{i,h} \in \Omega_i(h)$, rather have to be understood as *reduced procedures*. The explicit probability distribution associated with a reduced procedure, $\rho_{0,h'} \in \mathcal{P}_i(h)$, basically subsumes the probability distributions of procedures of all levels into one explicit distribution indirectly defined on $\mathcal{A}_i(h)$.

simple examples, i.e. Γ_1 and Γ_2, we have formalized the idea that players choose for procedures rather than actions. In this way we have separated choices for procedures and actual decisions. In the remainder of the paper we use this class of procedural games in order to isolate the impact of procedural choices on the strategic behavior of agents with belief-dependent utilities.

3 Procedural Games with Psychological Incentives

It is easy to see that if agents are only interested in their own expected material payoff, the set of all subgame perfect equilibria of two identical subgames is the same. Looking again at game Γ_3 in Figure 3, for example, this means that players are expected to react the same in histories h^1 and h^2. However, as already mentioned in the introduction, there exists ample evidence contradicting this traditional behavioral presumption. People very often react differently in outcomewise identical situations depending on the procedures which have led to them. Following the psychologist's view procedural choices affect peoples' beliefs about intentions. Hence, if people are (also) motivated by belief-dependent psychological payoffs, they exhibit procedural concerns. To conceptualize this idea, in this section we define procedural games in which agents have belief-dependent psychological incentives. This will allow us to formally capture the impact of procedural choices on the strategic behavior of emotional agents.

Economists have only recently developed a framework, i.e. psychological game theory, to formally account for behavioral traits such as emotions and intentions [e.g. Geanakoplos et al (1989) and Battigalli and Dufwenberg (2005)]. Psychological games are games in which agents are (also) motivated by belief-dependent psychological payoffs capturing their emotional involvement. Psychological payoffs arise from the beliefs that agents have about their opponents' strategies and beliefs. Therefore let agents have:

i) beliefs about the strategies of other players,

ii) beliefs about the beliefs of other players,

and

iii) let them update their beliefs as events unfold.

In order to formally capture assumptions *i*)-*iii*), we have to define an epistemic structure (*collectively coherent hierarchies of beliefs*) which describes what people initially belief and how they update their beliefs as play unfolds. This epistemic structure can be characterized in the context of our procedural games by assuming that players hold hierarchies of conditional beliefs over the procedural strategies as well as beliefs of other players $i \in \mathcal{N} \setminus \{0\}$.

As in Battigalli and Dufwenberg (2005) we only summarize the theory of hierarchies of conditional beliefs.[8] We describe, first, a system of conditional first-order-beliefs and then, secondly, show how this extends to higher orders (i.e. second-order-beliefs etc). In our class of procedural games denote by Ω_{-i} the set of procedural strategies of the opponents j where $j \in \mathcal{N} \setminus \{0, i\}$. At the beginning of any game, i.e. in the initial history h^0, player i does not know the true procedural strategies of his opponents. He only learns the true strategy $\omega_{-i} \in \Omega_{-i}$ step-by-step by updating his beliefs as the game unfolds. More formally, player i assigns probabilities to the events in the Borrel sigma algebra $\mathcal{B}$ of Ω_{-i} according to some probability measure. Let $\Delta(\Omega_{-i})$ be the set of all such probability measures. Denote $\mathcal{C} \subseteq \mathcal{B}$ the set of potential conditioning events at which player i can update his beliefs. In other words, $\mathcal{C}$ is the set of potentially observable events. Player i holds probabilistic beliefs about his opponents's procedural strategies conditional on each event $F \in \mathcal{C}$. These probabilistic beliefs are captured in a conditional probability system (*cps*).

From Battigalli Dufwenberg (2005) consider the following definition:

Definition 3 *A conditional probability system (cps) is a function* $\mu(\cdot|\cdot) : \mathcal{B} \times \mathcal{X} \to [0, 1]$ *defined on* $(X, \mathcal{B}, \mathcal{C})$ *such that for all* $E \in \mathcal{B}$ *and* $F', F \in \mathcal{C}$*:*

1. $\mu(\cdot|\cdot) \in \Delta(X)$,
2. $\mu(F|F) = 1$,
3. $E \subseteq F' \subseteq F$ *implies* $\mu(E|F) = \mu(E|F')\, \mu(F'|F)$,

where X *is a set, e.g.* Ω_{-i}*, whose 'true' element* $x \in X$ *is initially unknown and only learned step-by-step as conditioning events, e.g.* $F \in \mathcal{C}$*, are reached.*

Concentrating, first, on beliefs of order 1 means $X = \Omega_{-i}$. The first two conditions of definition 3 ensure that $\mu(\cdot|F)$ is indeed a probability measure (i.e. $\mu(\cdot|F) \in \Delta(X)$) which puts all probability weight on F given that F is observed. Condition 3 ensures that players update their beliefs according to Bayes' rule. The set of all functions μ for which conditions *1-3* hold is denoted by $\Delta^H(X)$. Hence, $\Delta^H(\Omega_{-i})$ is the set of all conditional probability systems of order 1 of player i.

Definition 3 can easily be extended to higher-order-beliefs. In the construction of the first-order *cps* we start from an initial situation in which player i does not know the true procedural strategy of his opponents. He has a conditional first-order-belief over it which is updated as play unfolds. Analog to this, in the construction of a second-order-belief we start from an initial situation in which player i does not know the true procedural strategy and the true conditional first-order-belief of players $-i$. Hence, the relevant set X in definition 3 becomes:

$$X = \Omega_{-i} \times \prod_{j \neq i} \Delta^H(\Omega_{-j}),$$

[8]For topological details, proofs and further references see Brandenburger and Dekel (1993) and Battigalli and Siniscalchi (1999).

where $i, j \in \mathcal{N} \backslash \{0\}$ and $\Delta^H(\Omega_{-j})$ is the set of conditional first-order *cps* of player j. The resulting conditional probability system does not only represent player i's belief about the strategy of players $-i$, but also about their first-order-beliefs.

Generalizing this idea, first- and higher-order *cps* are defined recursively as follows. Let:

$$\begin{aligned} X_{-i}^0 &= \Omega_{-i}, \text{ where } i \in \mathcal{N} \backslash \{0\}, \\ X_{-i}^k &= X_{-i}^{k-1} \times \prod_{j \neq i} \Delta^H\left(X_{-j}^{k-1}\right), \text{ where } i \in \mathcal{N} \backslash \{0\} \text{ and } k = 1, 2, \ldots . \end{aligned}$$

Then, a *cps* $\mu_i^k \in \Delta^H\left(X_{-i}^{k-1}\right)$ is called a k-order *cps* or simply a k-order belief. For $k > 1$, μ_i^k is a joint *cps* on the opponents' strategies and $(k-1)$-order *cps*', i.e.:

$$\begin{aligned} \mu_i^1 &\in \Delta^H\left(X_{-i}^0\right) \text{ where } X_{-i}^0 = \Omega_{-i}, \\ \mu_i^2 &\in \Delta^H\left(X_{-i}^1\right) \text{ where } X_{-i}^1 = \Omega_{-i} \times \Delta^H(\Omega_{-j}), \\ \mu_i^3 &\in \Delta^H\left(X_{-i}^2\right) \text{ where } X_{-i}^2 = \Omega_{-i} \times \Delta^H(\Omega_{-j}) \times \Delta^H\left(\Omega_{-j} \times \Delta^H(\Omega_{-i})\right) \text{ etc. } . \end{aligned}$$

This brings us to the formal definition of hierarchies of *cps*':[9]

Definition 4 *A hierarchy of cps is a countably infinite sequence of cps':*

$$\mu_i = \left(\mu_i^1, \mu_i^2, \ldots\right) \in \prod_{k>0} \Delta^H\left(X_{-i}^{k-1}\right).$$

As one can see, each piece of information appears many times in the belief hierarchy of player i. This implies that one can calculate marginal beliefs of higher-order-beliefs. As also Geanakoplos et al (1989) point out, these marginal beliefs of higher-order-beliefs should coincide with lower-order-beliefs in the belief hierarchy for the hierarchy to be meaningful. In other words beliefs should be *coherent*. We say a hierarchy of *cps*' is coherent if the *cps*' of distinct orders assign the same conditional probabilities to lower-order-events. This means,

$$\mu_i^k(\cdot|h) = marg_{X_{-i}^{k-1}} \mu_i^{k+1}(\cdot|h) \, (k = 1, 2, \ldots; h \in \mathcal{H}),$$

where $marg_{X_{-i}^{k-1}} \mu_i^{k+1}(\cdot|h)$ is the event of order $k-1$ in the *cps* of order $k+1$, $\mu_i^{k+1}(\cdot|h)$. If this condition holds, player i is said to have a coherent conditional belief system. It can be shown that a coherent hierarchy of *cps'* induces a single *cps* ν_i on the cross product of Ω_{-i} and the sets of hierarchies of *cps*' of i's opponents $-i$. Note, however, coherency regarding the own beliefs does not exclude the possibility that the *cps* ν_i puts a positive probability on the opponents *incoherence*. But as players are rational they should not believe that their opponents entertain incoherent beliefs. Hence, in order to rule this out, say that a coherent hierarchy μ_i satisfies belief in

[9] See also Battigalli and Dufwenberg (2005), p. 13.

coherency of order 1 if the induced *cps* ν_i is such that each $\nu_i(\cdot|h)$ with $h \in \mathcal{H}$ assigns probability one to the opponents' coherence of order 1. The hierarchy of coherent beliefs μ_i satisfies belief in coherency of order k, if it satisfies belief in coherency of order $k-1$, μ_i is *collectively coherent*, if it satisfies belief in coherency of order k for each positive integer k.[10] We denote the set of *collectively coherent hierarchies of beliefs* of player i by M_i. The set of collectively coherent beliefs of the opponents $-i$ is M_{-i} and $M = \prod_{j \in \mathcal{N} \setminus \{0\}} M_j$.

Finally, as the probability distributions associated with the moves of the player *chance*, i.e. player 0, are commonly known, nobody faces any uncertainty with regard to his *true type*. In other words, players do not learn the true strategy of player 0 over the course of the game, as it is ex ante commonly known. As will be seen in the Example presented below, this is crucial in the context of our procedural games. As the mixed strategy, ρ_0, of the player *chance* is commonly known, causal attribution is linked to procedural choices and not to outcomes.

To come to full circle, belief-dependent utilities are utilities that are not only defined on monetary outcomes, but also on collectively coherent hierarchies of beliefs and the commonly known probability distributions associated with the moves of the player *chance*:

Definition 5 *A belief dependent utility u of player i is a function:*

$$u_i : \mathcal{Z} \times \mathcal{P} \times \mathcal{M}_i \times \Pi_{j \neq i} (\Omega_j \times \mathcal{M}_j) \rightarrow \Re.$$

As mentioned in the introduction, strategic interactions with belief-dependent utilities have so far only been analyzed in traditional dynamic decision contexts, i.e. traditional extensive form representations, (e.g. Dufwenberg and Kirchsteiger (2004), Falk and Fischbacher (2006), Battigalli and Dufwenberg (2006)). Definition 5 represents an adaptation of these earlier approaches to our class of procedural games in which agents choose procedures rather than actions and strategies. In order to get a first impression of the impact of procedural choices on the interaction of emotional agents consider the following example:

Example: Assume that players 1 and 2 in game Γ_3 are reciprocal. This means they react kindly (unkindly) if they perceive the other to be kind (unkind). As we only want to give a first glimpse of the importance of procedures, we concentrate in this example on the perception that player 2 has about the kindness of player 1 in the histories h^1 and h^2. As said before, histories h^1 and h^2 are starting points of identical subgames.

Following Dufwenberg and Kirchsteiger (2004) the perceived kindness of player 2 in h^1 and h^2 can be defined as:

$$\lambda_{212} = \pi_2 \left(\mu_2^1 (\cdot|h^x), \mu_2^2 (\cdot|h^x), \rho_0 \right) - \pi_2^{e_1} \left(\mu_2^1 (\cdot|h^x), \mu_2^2 (\cdot|h^x), \rho_0 \right),$$

[10] See also Battigalli and Dufwenberg (2005), p.13.

where $x \in \{1,2\}$ and

$$\pi_2^{e_1}(\cdot) = \frac{1}{2}\Big[\max\left\{\pi_2\left(\mu_2^1(\cdot|h^x), \mu_2^2(\cdot|h^x), \rho_0\right), \omega_1 \in \Omega_1\right\} \\ + \min\left\{\pi_2\left(\mu_2^1(\cdot|h^x), \mu_2^2(\cdot|h^x), \rho_0\right), \omega_1 \in \Omega_1\right\}\Big].$$

The perceived kindness λ_{212} is defined as the difference between what player 2 believes player 1 intends to give him, $\pi_2(\cdot)$ (conditional on history h^x and given player 2's first- and second-order-beliefs, μ_2^1 and μ_2^2, and the mixed strategy of the player *chance*, ρ_0) and an equitable payoff, $\pi_2^{e_1}$. Dufwenberg and Kirchsteiger (2004) define the equitable payoff, $\pi_2^{e_1}$, as the average of the minimum and the maximum that player 2 believes player 1 could give him (again conditional on history h^x and given player 2's first- and second-order-beliefs, μ_2^1 and μ_2^2, and the mixed strategy of the player *chance*, ρ_0).

Assume, for example, that $\alpha_2 = (1 - \alpha_2) = \frac{1}{2}$ and imagine that player 2 believes that player 1 believes that he plays left in all the histories that he controls, i.e. histories h^1, h^2, h^3 and h^4. Given this, player 2 has to believe that player 1 intended to give him a material payoff of

$$\pi_2\left(\mu_2^1(\cdot|h^1), \mu_2^2(\cdot|h^1), \rho_0\right) = 200,$$

if he finds himself in history h^1 after player 1 has chosen procedure ω_{1,h^0}. In contrast to this, if player 2 finds himself in history h^2 he has to believe that player 1 intended to give him:

$$\pi_2\left(\mu_2^1(\cdot|h^2), \mu_2^2(\cdot|h^2), \rho_0\right) = \frac{1}{2}(200) + \frac{1}{2}(1800) = 1000,$$

by choosing procedure ω'_{1,h^0}. The equitable payoff, on the other hand, is given by:

$$\pi_2^{e_1}(\cdot) = \frac{1}{2}\Big[200 + 1800\Big],$$

where 200 is the minimum that player 2 believes player 1 could have given him in history h^0 (by playing ω_{1,h^0}) and 1800 is the maximum (by playing ω''_{1,h^0}). Putting the pieces together, player 2's perceived kindness in history h^1 and h^2 are respectively:

$$\lambda_{212}(h^1) = 200 - \frac{1}{2}\Big[200 + 1800\Big] = -800,$$
$$\lambda_{212}(h^2) = 1000 - \frac{1}{2}\Big[200 + 1800\Big] = 0.$$

Hence, although histories h^1 and h^2 are starting points of identical subgames, they are perceived very differently by player 2 due to the different procedural choices which have led to them. It is now easy to see that player 2 who is concerned about the intentions of player 1 might react differently in histories h^1 and h^2 depending on the strength of his reciprocal preferences. This gives a first idea of how procedural choices influence the causal attribution of responsibilities and the strategic interaction of emotional agents. ■

Given our class of procedural games as defined in the previous section and the belief-dependent utilities (Definition 5), we are now ready to define procedural games with psychological incentives:

Definition 6 *A procedural game with psychological incentives is a tuple:*

$$\Gamma_P = \left\langle \Gamma, (u_i)_{\mathcal{N}\setminus\{0\}} \right\rangle \text{ where } u_i : \mathcal{Z} \times \mathcal{P} \times \mathcal{M}_i \times \Pi_{j\neq i} (\Omega_j \times \mathcal{M}_j) \to \Re.$$

Procedural games with psychological incentives are the framework which we use to capture the impact of procedural choices on the interaction of psychologically motivated agents. Before presenting some applications, however, we subsequently adapt Battigalli and Dufwenberg (2005)'s *sequential equilibrium* to our class of procedural games with psychological incentives.

4 Sequential Psychological Equilibria in Procedural Games with Psychological Incentives

Battigalli and Dufwenberg (2005) adapt Kreps and Wilson (1982)'s concept of sequential equilibrium to their class of dynamic psychological games. They do so by characterizing *consistent assessments* that do not only consist of first-, but also of higher-order-beliefs and defining sequential equilibria as sequentially rational consistent assessments.

As in Battigalli and Dufwenberg (2005), also our equilibrium concept refers to mixed strategies, i.e. implicit randomizations over sets of procedures. Note, however, that, following Aumann and Brandenburger (1995), we interpret player i's mixed strategy as a conjecture on the part of his opponents as to what player i will do. Hence, denote a *behavioral procedural strategy* of player i as $\sigma_i = (\sigma_{i,h})_{h\in\mathcal{H}_i} \in \Sigma_i$, where Σ_i is the set of all mixed strategies of player i. The behavioral choice $\sigma_{i,h} \in \Sigma_i(h)$ in h has to be understood as an implicit randomization over the set of procedures $\Omega_i(h)$ in history h and interpreted as an array of common conditional first-order-beliefs held by i's opponents.[11] This means that the behavioral procedural strategy σ_i is part of an assessment $((\sigma, \rho_0), (\mu, \rho_0)) = ((\sigma_i, \rho_0), (\mu_i, \rho_0))_{i\in\mathcal{N}\setminus\{0\}}$ of behavioral strategies and hierarchies of conditional beliefs.

Three conditions ensure consistency of assessments in the original characterization by Kreps and Wilson (1982):

1. Beliefs must be derived using Bayes' rule,

2. Beliefs must reflect that players choose their strategies independently,

[11] See Battigalli and Dufwenberg (2005), p.16.

3. Players with identical information have identical beliefs.

In addition to these conditions, Battigalli and Dufwenberg (2005) add another requirement for consistency:

4. Players hold correct beliefs about each others beliefs.

Condition 1 holds by the definition of hierarchies of conditional belief systems (Definition 3). In other words, hierarchies of beliefs are defined in such a way that conditional beliefs are consistent with Bayes' rule. In order to formalize conditions 2-4 we first need to define what is meant by *stochastic independence*. Note, the observability of past actions allows us to define stochastic independence of the conditional belief systems in terms of marginal *cps*'. Different to the concept of marginal beliefs used in the previous section, a marginal *cps* now refers to player i's marginal belief on the procedural strategies of a particular player j and it is denoted as $\mu_{ij}^1 \in \Delta^H(\Omega_j)$, where $\Delta^H(\Omega_j)$ is the set of marginal *cps* on the procedural strategies of player j. Given this we can define stochastic independence of beliefs as:[12]

Definition 7 *A first-order cps $\mu_i^1 \in \Delta^H(\Omega_{-i})$ satisfies stochastic independence, if there exists a profile of marginal cps' $(\mu_{ij}^1)_{j\neq i} \in \prod_{j\neq i} \Delta^H(\Omega_j)$ such that $\mu_i^1(\omega_{-i}|h) = \prod_{j\neq i} \mu_{ij}^1(\omega_j|h)$ for all $h \in \mathcal{H}_i$. We denote the set of stochastically independent first-order cps' of a player i as $\Delta_I^H(\Omega_{-i})$.*

This brings us to our definition of consistent assessments:

Definition 8 *An assessment $((\sigma, \rho_0), (\mu, \rho_0))$ is consistent if:*

1. *The first-order cps of each player satisfies stochastic independence as formalized in Definition (7), i.e.:*

$$\forall i \in \mathcal{N} \setminus \{0\}, \mu_i^1 \in \Delta_I^H(\Omega_{-i}).$$

2. *The marginal first-order cps of any two players about any third player coincide, i.e.:*

$$\forall i \in \mathcal{N} \setminus \{0\}, \forall l \in \mathcal{N} \setminus \{i, j, 0\}, \forall h \in \mathcal{H}, \mu_{il}^1(\cdot|h) = \mu_{jl}^1(\cdot|h).$$

3. *Each players higher order beliefs in μ assign probability 1 to the lower order beliefs in μ itself:*

$$\forall i \in \mathcal{N} \setminus \{0\}, \forall k > 1, \forall h \in \mathcal{H}, \mu_i^k(\cdot|h) = \mu_i^{k-1}(\cdot|h) \times \delta_{\mu_{-i}^{k-1}},$$

where $\delta_{\mu_{-i}^{k-1}}$ is the probability measure which assigns probability 1 to μ_{-i}^{k-1}.

[12]See also Battigalli and Dufwenberg (2005)'s definition of stochastic independence, p. 17, and their definition of sequential equilibrium, p.19.

Conditions 1 and 2 capture the assumption that beliefs should be the end-product of a transparent reasoning process of intelligent people [Battigalli and Dufwenberg (2005)]. Condition 3, on the other hand, is analog to Geanakoplos et al (1989)'s condition requiring that players hold common and correct beliefs about each others' beliefs.

After having defined consistent assessments we can formally characterize *sequential psychological equilibria* (henceforth: SPE) by requiring sequential rationality:

Definition 9 *An assessment* $((\sigma, \rho_0), (\mu, \rho_0))$ *is a sequential psychological equilibrium (SPE), if for all* $i \in \mathcal{N} \setminus \{0\}, h \in \mathcal{H}_i$ *it holds:*

$$Supp\left(\sigma_{i,h}\right) \subseteq argmax_{\omega_{i,h} \in \Omega(h)} E_{\mu,\rho_0}\left[u_i | h, \omega_{i,h}\right],$$

where $E_{\mu,\rho_0}\left[u_i | h, \omega_{i,h}\right]$ *is the expected utility of player* i *conditional on history* h*, procedural choice* $\omega_{i,h} \in \Omega(h)$ *and given the system of hierarchies of conditional beliefs* μ *and the commonly known mixed strategy,* ρ_0*, played by player* 0*.*

Note, the expected utility of any player $i \in \mathcal{N} \setminus \{0\}$ (conditional on history h, *procedural choice* $\omega_{i,h} \in \Omega(h)$, given the system of consistent hierarchies of conditional beliefs μ and the commonly known mixed strategy, ρ_0) can be defined as:

$$\begin{aligned} E_{\mu,\rho_0}\left[u_i | h, \omega_{i,h}\right] = & \sum_{s_0 \in S_0(h)} \rho_0\left(s_0 | h\right) \sum_{\omega_{-i} \in \Omega_{-i}(h)} \mu_i^1\left(\omega_{-i} | h\right) \\ & \sum_{\omega_i \in \Omega_i\left(h, \omega_{i,h}\right)} \mu_{ji}^1\left(\omega_i | \left(h, \omega_{i,h}, \omega_{-i,h}\right)\right) u_i\left(\zeta\left(\omega_i, \omega_{-i}, s_0\right), \rho_0, \mu, \omega_{-i}\right), \end{aligned}$$

where $\zeta\left(\omega_i, \omega_{-i}, s_0\right) \in Z$ denotes the terminal history induced by the procedural strategies ω_i and ω_{-i}, and the strategy s_0 of player 0. Note, this specification is different from the expected utility formula traditionally used. Furthermore, it is also different from the specification used by Battigalli and Dufwenberg (2005) as it encloses the behavioral moves of the player *chance*.

The following proposition shows that there exists at least one sequential psychological equilibrium in any procedural game with psychological incentives and continuous utility functions:

Proposition 1 *If the utility functions are continuous, there exists at least one sequential psychological equilibrium assessment.*

Proof: Consider a procedural game with psychological incentives in which any procedure at any history is played with a strictly positive minimal probability ε. More formally, consider an ε-perturbed game Γ^ε in which players $i \in \mathcal{N} \setminus \{0\}$ dispose of 'constrained' choice sets $\Sigma_i^\varepsilon(h)$ at each history $h \in \mathcal{H}_i$. The 'constrained' choice set $\Sigma_i^\varepsilon(h)$ of player i in history h is defined as:

$$\Sigma_i^\varepsilon(h) := \{\tau_{i,h} \in \Sigma_i(h) | \tau_{i,h}(\omega_{i,h}) \geq \varepsilon, \forall \omega_{i,h} \in \Omega_i(h)\}.$$

So $\Sigma_i^\varepsilon(h)$ consists of only those elements in $\Sigma_i(h)$ that put a strictly positive probability greater or equal to ε on all elements $\omega_{i,h} \in \Omega_i(h)$, i.e. $\Sigma_i^\varepsilon(h) \subset \Sigma_i(h)$. It follows that in any Γ^ε the set of strictly mixed procedural strategies of players $i \in \mathcal{N}\backslash\{0\}$ is $\Sigma_i^\varepsilon = \times_{h\in\mathcal{H}_i} \Sigma_i{}^\varepsilon(h)$ and the set of all strictly positive behavioral procedural strategy profiles is $\Sigma^\varepsilon := \times_{i\in\mathcal{N}\backslash\{0\}} \Sigma_i^\varepsilon$. Note, for each $\sigma \in \Sigma^\varepsilon$ there exists a unique corresponding profile of hierarchies of *cps*' $\mu = \beta(\sigma)$ such that $((\sigma, \rho_0), (\beta(\sigma), \rho_0))$ is consistent.

Now, define for $\sigma \in \Sigma^\varepsilon$, $\varepsilon > 0$, $i \in \mathcal{N}\backslash\{0\}$ and $h \in \mathcal{H}_i$ the local best-response of player i in history h as:

$$BR_{i,h}^\varepsilon(\sigma) := \{\hat{\tau}_{i,h} \in \Sigma_i^\varepsilon(h) | u_i(\sigma_i/\hat{\tau}_{i,h}, \sigma_{-i}, \rho_0) \geq u_i(\sigma_i/\tau_{i,h}, \sigma_{-i}, \rho_0),\ \forall \tau_{i,h} \in \Sigma_i^\varepsilon(h)\},$$

where $\sigma_i/\tau_{i,h}$ denotes the behavioral procedural strategy for player i that specifies the strictly positive mixture $\tau_{i,h}$ at history $h \in \mathcal{H}_i$ and σ_i at every other history controlled by player i. In other words, local best-response-correspondences are strictly mixed behavioral choices that put at least a minimum probability ε on each procedure $\omega_{i,h} \in \Omega_i(h)$ given i's choices in all other histories controlled by him and given the behavioral procedural strategy of the opponents. The domain of the local best-response-correspondence is Σ^ε. The set $\Sigma^\varepsilon = \Sigma_1^\varepsilon \times \Sigma_2^\varepsilon ... \times \Sigma_N^\varepsilon$ and each Σ_i^ε with $i \in \mathcal{N}\backslash\{0\}$ is defined as $\Sigma_i^\varepsilon = \times_{h\in\mathcal{H}} \Sigma_i^\varepsilon(h)$. As said above, $\Sigma_i^\varepsilon(h)$ is the set of all behavioral procedural strategies of player i at history h that put at least a strictly positive probability ε on each procedure $\omega_{i,h} \in \Omega_i(h)$. It is non-empty (because $\Omega_i(h)$ is non-empty), compact and convex. Hence, also Σ^ε is non-empty, compact and convex (because the Cartesian product of nonempty, convex and compact sets is itself nonempty, convex and compact). Furthermore, $BR_{i,h}^\varepsilon(\sigma)$ is upper-semi-continuous. Note, the local best-response-correspondence $BR_{i,h}^\varepsilon(\sigma)$ is upper-semi-continuous, if for any sequence $(\sigma_i/\hat{\tau}_{i,h}^m, \sigma_{-i}^m) \rightarrow (\sigma_i/\hat{\tau}_{i,h}, \sigma_{-i})$ such that $\sigma_i/\hat{\tau}_{i,h}^m \in BR_{i,h}^\varepsilon(\sigma_i/\hat{\tau}_{i,h}^m, \sigma_{-i}^m)$ for all $m \in \{1, 2, ...\}$, we have $\sigma_i/\hat{\tau}_{i,h} \in BR_{i,h}^\varepsilon(\sigma_i/\hat{\tau}_{i,h}, \sigma_{-i})$. To see that this is indeed the case, note that for all m, the $u(\sigma_i/\hat{\tau}_{i,h}^m, \sigma_{-i}^m) \geq u(\sigma_i/\hat{\tau}_{i,h}', \sigma_{-i}^m)$ for all $\sigma_i/\hat{\tau}_{i,h}' \in \Sigma_i^\varepsilon$. Hence, by the continuity of the utility function, we have $u(\sigma_i/\hat{\tau}_{i,h}, \sigma_{-i}) \geq u(\sigma_i/\hat{\tau}_{i,h}', \sigma_{-i})$.

Given the local best-response correspondence $BR_{i,h}^\varepsilon(\sigma)$, the best-response correspondence $BR^\varepsilon(\sigma)$ is defined as:

$$BR^\varepsilon = (\hat{\tau}_{i,h})_{h\in\mathcal{H}_i \wedge i\in\mathcal{N}\{0\}}.$$

This implies that also $BR^\varepsilon : \Sigma^\varepsilon \rightarrow \Sigma^\varepsilon$ is upper semi continues, compact and convex and, hence, has a fixed point $\hat{\sigma}^\varepsilon$. As already pointed out by Geanakoplos et al (1989), the profile $\hat{\sigma}^\varepsilon$ constitutes an equilibrium of the constrained game Γ^ε.

Now, let ε^k be a sequence converging to 0 and $\hat{\sigma}^k$ the corresponding sequence of equilibrium assessments with $\hat{\sigma}^k$ being an equilibrium of Γ^{ε^k}. By the compactness of Σ, $\hat{\sigma}^k$ has an accumulation point σ^* and by the upper-semi-continuity of the local best-response-correspondents, $BR_h^\varepsilon(\sigma)$, $\sigma_{i,h}^*$ assigns positive probability only to those actions that are best responses to $(\sigma^*, \beta(\sigma^*), \rho_0)$ at h. Therefore $((\sigma^*, \rho_0), (\beta(\sigma^*), \rho_0))$ is a sequential equilibrium assessment. This concludes the proof. ■

Concluding, in this section we have formally defined sequential psychological equilibria in the context of our class of procedural games with psychological payoffs. Furthermore we have shown that every procedural game with psychologic incentives with continues utility functions has at least one SPE. Using our solution concept we demonstrate in the following section the impact of procedural choices on the interaction of psychologically motivated agents by means of two examples.

5 Applications

In the first application we analyze *a principal-agent relation* in which agents behave reciprocally towards their principal. This application shows the impact that different promotion procedures have on the interaction of psychologically motivated agents. In the second application we analyze the *'So long, Sucker'* game which has also been discussed by Dufwenberg and Kirchsteiger (2004). Different to them, however, we do not assume reciprocal behavior but guilt aversion. A full description of the strategic interaction with all possible sequential psychological equilibria is beyond the scope of this paper. We therefore limit the analysis to the characterization of only one equilibrium per application to demonstrate the impact and importance of procedural choices in the interaction of agents with belief-dependent utilities. Results and intuitions are presented in this section, lengthy mathematical proofs are relegated to the Appendix.

With the help of these two applications it is demonstrated i) how procedural choices influence the interaction of agents with belief-dependent psychological payoffs and ii) that the equilibrium predictions of the already existing literature using psychological games are sensitive to the availability of different procedures to take the same decision.

5.1 A Principal-Agent Relation

Imagine a principal, p, with two agents, $e1$ and $e2$, that is offered a project, b. He figures that in order to realize b he needs a *project manager*, pm, that is supported in the final phase of the project by an *assistant*, a, within the realm of his normal work. He knows that if both invest *high* effort, h, the project is a success, s, and he gets a payoff of $\pi(h,h) > 0$. If one of them invests *low* effort, l, however, the project will fail, f, and he will get a payoff of $\pi(h,l) = \pi(l,l) < \pi(h,h)$. Let both agents, $e1$ and $e2$, be equally skilled to perform either as *project manager* or *assistant*, implying that both have the same effort costs equal to v in case of *high* effort and 0 otherwise. Note that for simplicity we abstract in this principal-agent example from the usual question regarding the optimal incentive scheme. We take wages as given (e.g. due to a collective labor agreement) in order to single out the impact that the selection procedure has on the effort choices of the reciprocal agents. It is assumed that, in case of success, the principal pays $w(pm|s) > w(a)$ to the *project manager* and $w(a) < \frac{1}{2}((w(pm|s) - \upsilon) + (w(a) - \upsilon))$ to the *assistant*. On the other hand, in case of failure both get $w(pm|f) = w(a)$. Let efforts be observable, which implies that the *assistant* is aware of the *project managers*'s effort choice when choosing his own effort level, as he only collaborates in the final phase of the project. Furthermore, assume that the profits, $\pi(\cdot)$ minus the wage costs in case of a failure are 0 for the principal and positive if the project is a success.

Remark 1 *From the payment structure to the agents one can already see that, if effort is costly, the assistant has no monetary incentive to perform high effort since his wage will be* $w(a)$ *independent of the outcome of the project, b.*

The similarity of the two agents complicates the principal's decision on who is to become the *project manager* and who the *assistant*. Let the principal have two types of *procedures* that he can use to take his decision. He can either decide *behind closed doors*, *bcd*, or he can use a small *selection tournament*, *st*. This means his set of procedural strategies is $\Omega_p = \{st, bcd(e1), bcd(e2)\}$. For simplicity let the *selection tournament* be costless and credible to the agents. It is just about *concentration*, *c*, or *no concentration*, *nc*. Let it be commonly known that, if both *concentrate* or both do *not concentrate* during the short *selection tournament*, both are equally likely to become the *project manager*. If one *concentrates* and the other one does not, then the agent who *concentrates* gets the job.[13]

From the outset it is clear that the principal's profit is maximized if both his agents invest *high* effort and he shares part of the profit with the *project manager*. Against the background of Remark (1) it is easy to see, however, that if agents are only concerned about their own monetary payoff, it is impossible for the principal to elicit *high* effort from both agents after his selection decision.

Result 1 *If both agents are only concerned about their own monetary payoff, it is impossible for the principal to elicit high effort, h, from agents e1 and e2 independent of the selection procedure. As a consequence, the 'selection tournament' can never be strictly preferred to a decision 'behind closed doors'.*

Proof: As said in Remark (1), the *assistant* never has a monetary incentive to perform *high* effort (as it is costly) independent of the selection procedure. Obviously this is also known to the *project manager* who conjectures that no matter what he does the project will fail and he will get $w(pm|f) = w(a)$ independent of the selection procedure which the principal has used to take his decision. Hence, his optimal choice is also to always perform *low* effort, *l*. Given this the principal is indifferent between his two different types of selection procedures. ■

Consider now a situation in which agents $e1$ and $e2$ behave reciprocally towards the principal p. As pointed out before, this means they reciprocate kind with kind

[13]This means that if both perform equally during the *tournament* the principal flips a coin in front of them.

and unkind with unkind behavior. This type of behavior can be captured by assuming that each agent $i \in \{e1, e2\}$ maximizes the following utility function:

$$u_i = \pi_i + Y_{ip} \cdot (\kappa_{ip} \cdot \lambda_{ipi}),$$

where $Y_{ip} > 0$ is a positive constant that captures agents i's sensitivity to reciprocity, κ_{ip} is agent i's belief about his kindness towards the principal, λ_{ipi} is the agent i's perceived kindness of the principal towards himself and π_i is agent i's own expected monetary payoff. Note, this conceptualization of reciprocity is analog to the definition by Dufwenberg and Kirchsteiger (2004).

In the Example in section 3 we have already defined perceived kindness. For completeness, however, let us restate it here in the context of our *principal-agent relation*. Agent i's perceived kindness of the principal, λ_{ipi}, in history h^x is defined as:

$$\lambda_{ipi} = \pi_i \left(\mu_i^1 (\cdot|h^x), \mu_i^2 (\cdot|h^x), \rho_0\right) - \pi_i^{e_p} \left(\mu_i^1 (\cdot|h^x), \mu_i^2 (\cdot|h^x), \rho_0\right).$$

As before, $\pi_i(\cdot)$ describes what agent i believes the principal intends to give him and $\pi_i^{e_p}(\cdot)$ is the equitable payoff which characterizes agent i's belief about the average that the principal could have given him. More formally:

$$\begin{aligned}\pi_i^{e_p}(\cdot) =& \frac{1}{2}\Big[\max\left\{\pi_i \left(\mu_i^1 (\cdot|h^x), \mu_i^2 (\cdot|h^x), \rho_0\right), \omega_p \in \{st, bcd(e1), bcd(e2)\}\right\} \\ &+ \min\left\{\pi_i \left(\mu_i^1 (\cdot|h^x), \mu_i^2 (\cdot|h^x), \rho_0\right), \omega_p \in \{st, bcd(e1), bcd(e2)\}\right\}\Big].\end{aligned}$$

Similarly agent i's kindness towards the principal in history h^x can be described as:

$$\kappa_{ip} = \pi_p \left(\mu_i^1 (\cdot|h^x), \omega_i, \rho_0\right) - \pi_p^{e_i} \left(\mu_i^1 (\cdot|h^x), \omega_i, \rho_0\right),$$

where,

$$\begin{aligned}\pi_p^{e_i}(\cdot) =& \frac{1}{2}\Big[\max\left\{\pi_i \left(\mu_i^1 (\cdot|h^x), \omega_i, \rho_0\right), \omega_i \in \{l, h\}\right\} \\ &+ \min\left\{\pi_i \left(\mu_i^1 (\cdot|h^x), \omega_i, \rho_0\right), \omega_i \in \{l, h\}\right\}\Big].\end{aligned}$$

In line with the above, the expected material payoff $\pi_p(\cdot)$ describes what agent i believes the principal will get, given his beliefs, the commonly known 'mixed strategy' of player *chance* and his own choice $\omega_i \in \{l, h\}$. Furthermore, $\pi_p^{e_i}(\cdot)$ is agent i's belief about the the average that he can give to the principal p.

In contrast to Result 1, the question arises whether the profit maximizing principal is also indifferent between his selection procedures, given that the agents behave reciprocally towards him. Note that the *principal-agent relation* is symmetric. This allows us to state the following result in terms of *project manager* and *assistant* rather than the behavior of agents $e1$ and $e2$ in their different possible roles.

Result 2 *If the project manager's sensitivity to reciprocity is:*

$$Y_{pmp} \geq \frac{(w(pm|f) - w(pm|s)) + \upsilon}{\frac{1}{2}\left[\frac{1}{2}\left[w\left(pm|s\right) - w(a)\right] - \upsilon\right]\left[\Delta\pi_p + \Delta w(pm)\right]},$$

and the assistant's sensitivity to reciprocity is:

$$Y_a \geq \frac{\upsilon}{\frac{1}{2}\left[\frac{1}{2}\left[w\left(pm|s\right) - w(a)\right] - \upsilon\right]\left[\Delta\pi_p + \Delta w(pm)\right]},$$

where $\Delta\pi_p = \pi_p\left(h, h\right) - \pi_p\left(h, l\right)$ *and* $\Delta w(pm) = w(pm|f) - w(pm|s)$*, then the sequential psychological equilibrium is given by:*

1. *The project manager i) chooses low effort following a decision of the principal taken behind closed doors, ii) chooses concentration and iii) high effort, if the principal uses a selection tournament to take his decision.*

2. *The assistant i) chooses low effort following a decision of the principal taken behind closed doors, ii) chooses concentration, iii) low effort following low effort by the project manager and the selection tournament and iv) high effort, if the principal uses a selection tournament to take his decision and the project manager has chosen high effort as well.*

3. *The principal uses the selection tournament.*

Proof: See appendix

The intuition behind this result is the following: The *assistant* feels unkindly treated, if the principal has taken the decision *behind closed doors*.[14] As effort is costly, he thus chooses *low* effort independent of the effort choice of the *project manager*. In comparison to that, the *assistant* does feel kindly treated if the principal has used a *selection tournament* to choose the *project manager*. Thus, if he is sensitive enough to reciprocity, i.e. condition Y_a in Result (2) holds, then he chooses *high* effort given that the *project manager* has chosen *high* effort following *st*. But if, on the other hand, the *project manager* has chosen *low* effort following the *selection tournament*, the *assistant* knows that it is useless to invest *high* effort and, hence, he optimally chooses *l*. The *project manager* obviously knows all this. Hence, if he was selected *behind closed doors*, he chooses *low* effort because he knows that the *assistant* will. If he was selected via a *selection tournament*, however, and he knows that the assistant is sufficiently sensitive to reciprocity he will choose *high* effort to reciprocate the kind behavior of the principal. Given this the principal will choose the *selection tournament*, as in this way he maximizes his own profit. This highlights the importance of procedural choices in the interaction of psychologically motivated agents.

[14]This is also in analogy to the 'promotion' example briefly sketched in the introduction.

In addition, one can also confront Result (2) with the results obtained in the setting of Dufwenberg and Kirchsteiger (2004) who do not allow for different types of procedures. In order to do so, consider the same situation as described above, but without the principal's possibility to perform a *selection tournament*. In other words, the principal can only decide *behind closed doors*.

Result 3 *For all $Y_{ap} \geq 0$ and $Y_{pmp} \geq 0$ the SPE is given by:*

1. *The assistant always chooses low effort either out of pure cost (if $Y_{ap} = 0$) or cost and kindness considerations (if $Y_{ap} > 0$).*

2. *The project manager knows this and, consequently, also chooses low effort independent of Y_{pmp}.*

3. *The principal is indifferent between choosing agent $e1$ or $e2$ as project manager. Hence, any choice of the principal is part of an equilibrium.*

Proof: The *assistant* will always perceive the principal's decision as unkind. Hence he is never inclined to choose *high* effort out of kindness considerations. This is even reinforced by the fact that *high* effort is costly. Consequently the *assistant*'s optimal strategy is to choose l in every history in which he is active. As said above, the *project manager* knows this and figures that what ever he does the project will fail. Hence, his optimal choice is also to invest *low* effort. Given this the principal is indifferent between $bcd(e1)$ and $bcd(e2)$. ■

As one can see, if alternative procedures to take the same decision are neglected different equilibrium predictions result. This is not a mere artifact in this particular example but holds true also in other settings as will also be seen in the next application. Hence, the behavioral predictions that have been made in the hitherto existing literature on psychological games [e.g. Dufwenberg and Kirchsteiger (2004) and Falk and Fischbacher (2006)] are sensitive to the availability of different procedures to take the same decision.

Concluding, we have seen in this *principal-agent* example how procedural concerns inherently arise if agents are also psychologically motivated. Furthermore, taking different procedures to take the same decision explicitly into account leads to behavioral predictions that differ from the results with mere consequentialist preferences, as traditionally assumed in economic theory, and they also differ from the results obtained in settings allowing for belief-dependent utilities but neglecting procedural choices. In the next application we will demonstrate the impact of procedural choices when agents are guilt averse [e.g. Charness and Dufwenberg (2006), Battigalli and Dufwenberg (2006)].

5.2 The 'So long, Sucker' Game with Guilt Aversion

Consider the game in Figure 4:[15]

[Figure 4 here]

This *'So long, Sucker' game* is a three-player game in which a player 1 seems to be trapped since he has to be unkind to one of the other players 2 and 3. This setting has already been analyzed by Nalebuff and Shubik (1988) and Dufwenberg and Kirchsteiger (2004) to explain why agents might choose to punish others (in this case player 1), i.e. reciprocate for any perceived unkindness, even if it is costly for themselves.

Different to Nalebuff and Shubik (1988) as well as Dufwenberg and Kirchsteiger (2004) assume that agent 1 is guilt averse. More precisely, assume that player 1 feels guilty, if the other two players get the impression that he did not treat them equally. This can be conceptualized as follows: At any endnode $z \in Z$ player j's inference ($j \in \{2,3\}$) with regard to what player 1 intended to give him by playing the procedural strategy ω_i is:

$$E_{\mu_j^1,\mu_j^2,p_0}\left[\pi_j|\mathcal{H}\left(z\right),\omega_1\right].$$

Obviously, player j also has a belief in z about what player 1 intended to give to the other player q, where $q \neq j \wedge q \neq 1$:

$$E_{\mu_j^1,\mu_j^2,p_0}\left[\pi_q|\mathcal{H}\left(z\right),\omega_1\right].$$

This means player j can infer player 1's intended difference, i.e. player 1's favoritism, between j and the other player q:

$$E_{\mu_j^1,\mu_j^2,p_0}\left[\pi_j|\mathcal{H}\left(z\right),\omega_1\right]-E_{\mu_j^1,\mu_j^2,p_0}\left[\pi_q|\mathcal{H}\left(z\right),\omega_1\right].$$

In line with the above-sketched intuition concerning player 1's guilt feeling and similar to Battigalli and Dufwenberg (2006), we say that player 1 is affected by 'guilt from blame', if players 2 and 3 get a perception of favoritism. His preferences can hence be written as:

$$u_1\left(z,\mu_{-1}^1,\mu_{-1}^2\right)=\pi_1-\sum_j Y_{1j}\left(\left|E_{\mu_j^1,\mu_j^2,p_0}\left[\pi_j|\mathcal{H}\left(z\right),\omega_1\right]-E_{\mu_j^1,\mu_j^2,p_0}\left[\pi_q|\mathcal{H}\left(z\right),\omega_1\right]\right|\right),$$

where Y_{1j} is a positive constant capturing player 1's sensitivity to guilt and μ_{-1}^1 and μ_{-1}^2 are the other players first- and second-order-beliefs. Note, in each history player 1 maximizes his utility u_i conditional on his belief up to the third-order because he takes his belief about the other players' second-order-belief, μ_{-1}^2, into account.[16] For

[15]Note that actions that are played by player *chance* with probability 0 are disregarded in Figure 4.

[16]For comparison see Battigalli and Dufwenberg (2006)'s definition of 'guilt from blame'.

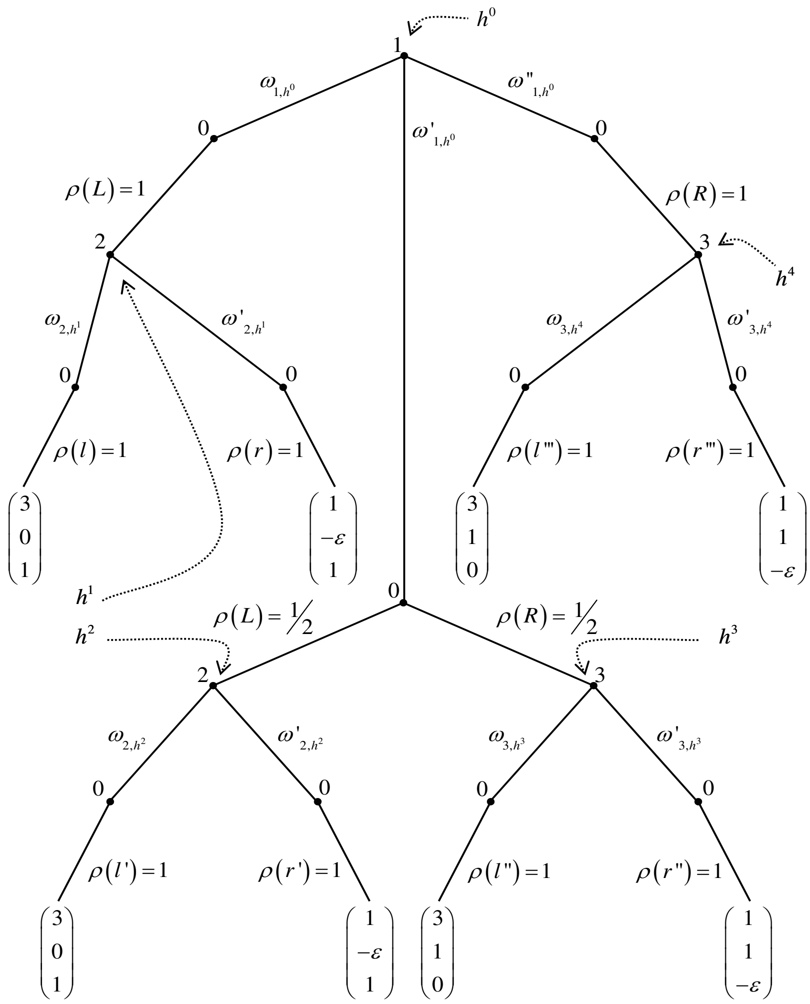

Figure 4: 'So long, Sucker' Game

simplicity, assume that players 2 and 3 perceive player 1's favoritism, but this does not have any effect on their utility.[17] In other words, players 2 and 3 are only concerned about their own material welfare.

As a benchmark let us first state how player 1 behaves if all players are only concerned about his own monetary payoff:

Result 4 *For all $\varepsilon \geq 0$, if all players are only interested in their own material payoff, then player 1 is indifferent with regard to his procedural choice.*

Proof: By backward induction, players 2 and 3 respectively choose $\{\omega_{2,h^1}, \omega_{2,h^2}\}$ and $\{\omega_{3,h^3}, \omega_{3,h^4}\}$ in the histories that they control. This implies player 1 knows that he gets 3 for sure independent of his own choice. Hence, he is indifferent between ω_{1,h^0}, ω'_{1,h^0} and ω''_{1,h^0}. ■

The situation changes assuming that player 1 is guilt averse as defined above. Given our set up with guilt aversion, it is possible to state the following result:

Result 5 *If $Y_{12} > 0$ and $Y_{13} > 0$, then the only SPE is given by:*

1. *Player 1 chooses ω'_{1,h^0} in history h^0.*
2. *Players 2 and 3 choose respectively $\{\omega_{2,h^1}, \omega_{2,h^2}\}$ and $\{\omega_{3,h^3}, \omega_{3,h^4}\}$ in the histories that they control.*

Proof: In line with player 2's and 3's preferences, let player 1's first-order-belief and player 1's belief about the second-order-belief of players 2 and 3 be $\{\omega_{2,h^1}, \omega_{2,h^2}\}$ and $\{\omega_{3,h^3}, \omega_{3,h^4}\}$. This implies that player 1's belief about players 2's and 3's perception of his intended favoritism is:

i) $(0 - 1) = -1$ (player 2) and $(1 - 0) = 1$ (player 3), if he chooses ω_{1,h^0},

ii) $\frac{1}{2}(0 - 1) + \frac{1}{2}(1 - 0) = 0$ (for both players), if he chooses ω'_{1,h^0} and

iii) $(1 - 0) = 1$ (player 2) and $(0 - 1) = -1$ (player 3), if he chooses ω''_{1,h^0}.

This means, his guilt feeling is minimized by playing ω'_{1,h^0}. Furthermore, his own expected material payoff given his first-order-beliefs is 3 independent of his procedural choice. Therefore, it is easy to see that the rational player 1 that is guilt averse optimally chooses the procedure ω'_{1,h^0} to take his decision between players 2 and 3. In addition, players 2 and 3 choose $\{\omega_{2,h^1}, \omega_{2,h^2}\}$ and $\{\omega_{3,h^3}, \omega_{3,h^4}\}$ in line with player 1's beliefs because of their material concerns. This concludes the proof. ■

Note, also in this example it holds that procedures mitigate the own as well as the others' psychological payoffs.

[17] Note that one could in addition assume that players 2 and 3 are disappointed due to the perceived favoritism. This would, however, only complicate the analysis without changing the results.

Remark 2 *As in the previous example, ignoring player 1's possibility to choose a randomization procedure to take his decision, i.e.* ω'_{1,h^0}*, would lead to a different behavioral prediction. He would be indifferent between choosing* ω_{1,h^0} *and* ω''_{1,h^0}*.*

Hence, also here it holds that neglecting different procedures to take the same decision, as it is done in the hitherto existing literature on psychological games, leads to different equilibrium predictions. This highlights again, how procedural concerns can be conceptualized as an inherent part of the interaction of agents with belief-dependent utilities.

All in all, in this section we have used the concept of sequential psychological equilibrium developed in the previous section to formally demonstrate the impact of procedural choices on the strategic interaction of emotional agents. We have seen how procedural choices influence their interactions and how the inclusion of different procedures to take the same decision affects the behavioral predictions of the existing literature on psychological games.

6 Conclusion

Any decision in human interactions is inherently associated with a procedure which characterizes the way in which the decision is taken. This means it is impossible to take a decision without deciding first on *how* to take it. It is widely accepted in other scientific disciplines and it has been shown experimentally that people react differently in outcomewise identical situations depending on the procedures which have led to them. People are concerned about the way in which decisions are taken. Nevertheless traditional economic theory has neglected the impact of procedural choices on human interaction. It has ignored the influence of procedures on human interactions as traditional economic theory is based on consequentialist preferences which are difficult to reconcile with the existing evidence on procedural concerns.

Only in recent years psychological game theory has contested the consequentialist view in economic theory by assuming that agents also sense psychological payoffs which, broadly speaking, depend on agents' beliefs about the other's strategies and beliefs. It has been shown in our paper how procedural concerns can be conceptualized in a game theoretic setting assuming that agents are (also) incentivized by belief-dependent psychological payoffs. According to our approach procedural choices influence the beliefs that people hold with regard to others. In this way they mitigate the causal attribution of responsibilities and the evaluation of intentions.

With the help of two applications we have furthermore demonstrated i) how procedural concerns influence the strategic interaction of agents with belief-dependent utilities and ii) that the equilibrium predictions in the already existing literature on psychological games are sensitive to the availability of different procedures to take the same decision. The hitherto existing literature on psychological games solely concentrates on situations in which agents are held fully responsible for all consequences of their actions. In contrast to this, in our class of procedural games agents can choose between different procedures. They can influence the process of causal attribution and the evaluation of intentions. Consequently, different equilibrium predictions arise.

Concluding, procedural concerns can play an important role in areas of eminent concern to economists. Hence, they should not be neglected.

7 References

1. Anderson, R. and Otto, A. (2003), *Perceptions of fairness in the justice system: A cross-cultural comparison*, Social Behavior and Personality, 31, 557-564.

2. Aumann, R. and Brandenburger, A. (1995), *Epistemic conditions for Nash equilibrium*, Econometrica, 63, 1161-1180.

3. Battigalli, P. and Siniscalchi, M. (1999), *Hierarchies of Conditional Beliefs and Interactive Epistemology in Dynamic Games*, Journal of Economic Theory, 88, 188-230.

4. Battigalli, P. and Dufwenberg, M. (2005), *Dynamic Psychological Games*, Mimeo, IGIER-Bocconi.

5. Battigalli, P. and Dufwenberg, M. (2006), *Guilt in Games*, American Economic Review, Papers and Proceedings, forthcoming.

6. Bies, R. and Tyler, T. (1993), *The 'litigation mentality' in organizations: A test of alternative psychological explanations*, Organization Science, 4, 352-366.

7. Blader, S. and Tyler, T. (2003), *A four-component model of procedural justice: Defining the meaning of a 'fair' process*, Personality and Social Psychology Bulletin, 29, 747-758.

8. Blount, S. (1995), *When Social Outcomes Aren't Fair: The Effect of Casual Attributions on Preferences*, Organizational Behavior and Human Decision Processes, 63, 131-144.

9. Bohnet, I. (2006), *How Institutions Affect Behavior: Insights from Economics and Psychology*, in: De Cremer, D., Zeelenberg, M. and Murnighan, K. (eds), *Social Psychology and Economics*, London: Lawrence Erlbaum, 2006, 213-238.

10. Bohnet, I. (2007), *Institutions and Trust: Implications for Preferences, Beliefs and Behavior*, Rationality and Society, 19(1), 99-135.

11. Bohnet, I. and Zeckhauser, R. (2004), *Trust, Risk and Betrayal*, Journal of Economic Behavior and Organization, 55, 467-484.

12. Bolton, G. and Ockenfels, A. (2000), *ERC: A Theory of Equity, Reciprocity, and Competition*, American Economic Review, 90(1), 166-193.

13. Bolton, G. Brandts, J. and Ockenfels, A. (2005), *Fair Procedures: Evidence from Games Involving Lotteries*, Economic Journal, 115(506), 1054-1076.

14. Bowles, S. (1998), *Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions*, Journal of Economic Literature, 36(1), 75-111.

15. Brandenburger and Dekel (1993), *Hierarchies of Beliefs and Common Knowledge*, Journal of Economic Theory, Elsevier, 59(1), 189-198.

16. Brandts, J., Gueth, W. and Stiehler, A. (2006), *I want you!: An experiment studying the selection effect when assigning distributive power*, Journal of Labor Economics, 13, 1-17.

17. Charness, G. and Rabin, M. (2002), *Understanding Social Preferences With Simple Tests*, Quarterly Journal of Economics, 117(3), 817-869.

18. Charness, G. (2004), *Attribution and Reciprocity in an Experimental Labor Market*, Journal of Labor Economics, 22(3), 553-584.

19. Charness, G. and Dufwenberg, M. (2006), *Promises and Partnership*, Econometrica, 74(6), 1579-1601.

20. Charness, G. and Levine, D. (2007), *Intention and Stochastic Outcomes: An Experimental Study*, Economic Journal, 117(522), 1051-1072.

21. Collie, T., Bradley, G. and Sparks, B. (2002), *Fair process revisited: Differential effects of interactional and procedural justice in the presence of social comparison information*, Journal of Experimental Social Psychology, 38, 545-555.

22. De Cremer, D. and van Knippenberg, D. (2003), *Cooperation with leaders in social dilemmas: On the effects of procedural fairness and outcome favorability in structural cooperation*, Organizational Behavior and Human Decision Processes, 91, 1-11.

23. Dufwenberg, M. and Kirchsteiger, G. (2004), *A Theory of Sequential Reciprocity*, Games and Economic Behavior, 47(2), 268-298.

24. Elster, J. (1998), *Emotions and Economic Theory*, Journal of Economic Literature, 36(1), 47-74.

25. Falk, A. and Fischbacher, U. (2006) *A Theory of Reciprocity*, Games and Economic Behavior, 54(2), 293-315.

26. Falk, A., Fehr, E. and Fischbacher, U. (2000), *Testing Theories of Fairness - Intentions Matter*, Working Paper No. 63, University of Zurich, forthcoming in: Games and Economic Behavior.

27. Fehr, E. and Schmidt, K. (1999), *A Theory Of Fairness, Competition, And Cooperation*, The Quarterly Journal of Economics, 114(3), 817-868.

28. Fehr, E. and Gaechter, S. (2000), *Fairness and Retaliation: The Economics of Reciprocity*, Journal of Economic Perspectives, 14(3), 159-181.

29. Geanakoplos, J., Pearce, D. and Stacchetti, E. (1989) *Psychological Games and Sequential Rationality*, Games and Economic Behavior, 1, 60-79.

30. Heider, F. (1958), *The Psychology of Interpersonal Relations*, New York: John Wiley & Sons.

31. Hong, K. and Bohnet, I. (2005), *Status and Distrust: The Relevance of Inequality and Betrayal Aversion*, KSG Working Paper No. RWP04-041, Harvard PON Working Paper.

32. Kelley, H. (1967), *Attribution in social psychology*, Nebraska Symposium on Motivation, 15, 192-238.

33. Kelley, H. (1973), *The processes of causal attribution*, American Psychologist, 28, 107-128.

34. Konovsky, M. (2000), *Understanding Procedural Justice and Its Impact on Business Organizations*, Journal of Management, 26(3), 489-511.

35. Krawczyk, M. (2007), *A model of procedural and distributive fairness*, Mimeo, University of Amsterdam.

36. Kreps, D. and Wilson, R. (1982), *Sequential Equilibria*, Econometrica, 50(4), 863-894.

37. Lemons, M. and Jones, C. (2001), *Procedural justice in promotion decisions: using perceptions of fairness to build employee commitment*, Journal of Managerial Psychology, 16(4), 268-281.

38. Lind, E. and Tyler, T. (1988), *The social psychology of procedural justice*, New York, NY, US: Plenum Press.

39. Lind, E., Greenberg, J., Scott, K. and Welchans, T. (2000), *The winding road from employee to complainant: Situational and psychological determinants of wrongful termination claims*, Administrative Science Quarterly, 45, 557-590.

40. Murphy, K. (2004), *The role of trust in nurturing compliance: A study of accused tax avoiders*, Law and Human Behavior, 28, 187-209.

41. Nalebuff, B. and Shubik, M. (1988) *Revenge And Rational Play*, Papers 138. Princeton. Woodrow Wilson School - Public and International Affairs.

42. North, D. (1991), *Institutions*, Journal of Economic Perspective, 5(1), 97-221.

43. Sebald, A. (2007), *Procedural Concerns and Reciprocity*, ECORE Discussion Paper, 2007/54, Bruxelles.

44. Thibaut J. and Walker, L. (1975), *Procedural Justice*, Erlbaum, Hillsdale, NJ.

45. Trautmann, S. (2006), *A Fehr-Schmidt Model for Process Fairness*, Working Paper, CREED, University of Amsterdam.

46. Tyler, T. (1990), *Why people obey the law*, New Haven, CT, US: Yale University Press.

47. Tyler, T. (2003), *Procedural justice, legitimacy, and the effective rule of law*, in Tonry, M. (ed.), Crime and justice: A review of research, 30, 283-358.

48. Rabin, M. (1993), *Incorporating Fairness into Game Theory and Economics*, American Economic Review, 83(5), 1281-1302.

49. Roberts, K. and Markel, K. (2001), *Claiming in the name of fairness: Organizational justice and the decision to file for workplace injury compensation*, Journal of Occupational Health Psychology, 6, 332-347.

50. Ross, M. and Fletcher, G. (1985), *Attribution and Social Perception*, in G. Lindsey & E. Aronson (eds.). The Handbook of Social Psychology, 2, 73-114.

51. Wenzel, M. (2002), *The impact of outcome orientation and justice concerns on tax compliance: The role of taxpayers' identity*, Journal of Applied Psychology, 87, 629-645.

8 Appendix

8.1 Proof of Result 2

Note, as the *principal-agent relation* is symmetric we will concentrate on the behavior of *project manager* and *assistant* rather than the behavior of agents $e1$ and $e2$ in the different possible roles. Let us start by looking at the behavior of the *assistant* following the decision by the principal to take the decision *behind closed doors*. Remember, when he has to decide about his effort level, he knows about the *project manager*'s effort level, the principal's procedural choice etc, i.e. he is perfectly informed about the history of the game he is in.

To start with, assume that in any history that the *assistant* can find himself following *bcd* he believes that the principal believes, i.e. the *assistant* 's second-order-belief, that:

1. the *assistant* chooses *low* effort, given that the principal has taken the decision *behind closed doors*,

2. the *project manager* and the *assistant* will choose *concentration* and *high* effort, given that the principal has taken the decision by means of a *selection tournament*.

This means, if it is the *assistant*'s turn and the principal has taken his decision *behind closed doors*, then the *assistant* believes that the principal intends to give him:

$$\pi_a(\cdot) = w\left(a\right). \tag{1}$$

Given this and the *assistant*'s second order belief his perceived kindness of the principal following *bcd* is:

$$\lambda_{apa} = w\left(a\right) - \frac{1}{2}\left(w(a) + \frac{1}{2}\left(\left(w(a) - \upsilon\right) + \left(w(pm|s) - \upsilon\right)\right)\right) < 0, \tag{2}$$

where $\frac{1}{2}\left(w(a) + \frac{1}{2}\left(\left(w(a) - \upsilon\right) + \left(w(pm|s) - \upsilon\right)\right)\right)$ is the *assistant*'s belief about the average that the principal could have given him and $\frac{1}{2}\left(\left(w(a) - \upsilon\right) + \left(w(pm|s) - \upsilon\right)\right)$ is the *assistant*'s expected payoff given that the principal had chosen the *selection tournament*. Furthermore, the *assistant*'s kindness is either:

$$\kappa_{ap}(l) = \left(\pi_p\left(l, l\right) - w\right) - \frac{1}{2}\left(\left(\pi_p\left(l, h\right) - w\right) + \left(\pi_p\left(l, l\right) - w\right)\right) = 0, \tag{3}$$

if he chooses *low* effort or

$$\kappa_{ap}(h) = \left(\pi_p\left(l, h\right) - w\right) - \frac{1}{2}\left(\left(\pi_p\left(l, h\right) - w\right) + \left(\pi_p\left(l, l\right) - w\right)\right) = 0, \tag{4}$$

if he chooses *high* effort, conditional on the *low* effort by the *project manager*. Conditional on the *high* effort by the *project manager* his kindness towards the principal is either:

$$\kappa_{ap}(l) = \left(\pi_p\left(h, l\right) - w\right) - \frac{1}{2}\left(\left(\pi_p\left(h, h\right) - w\right) + \left(\pi_p\left(h, l\right) - w\right)\right) > 0, \tag{5}$$

if he chooses *low* effort or

$$\kappa_{ap}(h) = \left(\pi_p\left(h, h\right) - w\right) - \frac{1}{2}\left(\left(\pi_p\left(h, h\right) - w\right) + \left(\pi_p\left(h, l\right) - w\right)\right) < 0, \tag{6}$$

if he chooses *high* effort. To summarize the perceptions and the optimal behavior of the *assistant*:

1. If the principal chooses to take the decision *behind closed doors* and given the *assistant*'s aforementioned second-order-beliefs, the perceived kindness of the *assistant* is negative independent of what the *project manager* does.

2. The *assistant*'s kindness towards the principal can: i) be 0 independent of his own choice, if the *pm* chooses low effort as well, or, ii) positive and negative given that the *project manager* has chosen *high* effort.

3. Hence, first, as effort is costly, he optimally chooses *low* effort, given *low* effort by the *project manager*. Secondly, as effort is costly and perceived kindness is negative, he also optimally chooses *low* effort given *high* effort by the *project manager*. Note, the *assistant*'s optimal behavior is in line with his second order beliefs.

In contrast to this, the *assistant*'s perceived kindness of the principal following the *selection tournament* is:

$$\begin{aligned}\lambda_{apa} &= \frac{1}{2}\left(w\left(a\right)-\upsilon\right)+\frac{1}{2}\left(w(pm|s)-\upsilon\right) \\ &- \frac{1}{2}\left(w(a)+\frac{1}{2}\left(\left(w(a)-\upsilon\right)+\left(w(pm|s)-\upsilon\right)\right)\right)>0,\end{aligned} \tag{7}$$

where $\frac{1}{2}\left(w\left(a\right)-\upsilon\right)+\frac{1}{2}\left(w(pm|s)-\upsilon\right)$ is the *assistant*'s belief about what the principal intended to give him by choosing the *selection tournament*. Note, it can easily be seen that the *assistant*'s kindness towards the principal in the histories in which he is active is the same as under *bcd*, i.e. equations (3),(4),(5) and (6). From equation (7) we already see that the *assistant* perceives the *selection tournament* as kind. Given this the question arises whether and under what conditions this would make him choose *high* effort. Rationality requires that he chooses *high* effort if his utility from choosing *high* effort is bigger or equal to his utility from choosing *low* effort, i.e.:

$$u_a(h) \geq u_a(l), \tag{8}$$

which means

$$(w(a)-\upsilon)+Y_{ap}\left(\kappa_{ap}(h)\lambda_{apa}\right) \geq w(a)+Y_{ap}\left(\kappa_{ap}(l)\lambda_{apa}\right). \tag{9}$$

As $\kappa_{ap}(l)$ and $\kappa_{ap}(h)$ are 0 in histories following *st* and *low* effort by the *pm*, it can easily be seen that equation (9) never holds as $\upsilon > 0$. Hence, the *assistant* always chooses *low* effort under the *selection procedure* given that the *project manager pm* has chosen *low* effort as well. In case the *project manager* has chosen *high* effort, however, the situation changes. Equation (9) can be rewritten as:

$$Y_{ap} \geq \frac{\upsilon}{\lambda_{apa}\left(\kappa_{ap}(h)-\kappa_{ap}(l)\right)}. \tag{10}$$

Plugging in for λ_{apa} and $\kappa_{ap}(\cdot)$ gives:

$$\begin{aligned}Y_{ap} &\geq \frac{\upsilon}{\frac{1}{2}\left[\frac{1}{2}\left[w\left(pm|s\right)-w(a)\right]-\upsilon\right]\left[\pi_p\left(h,h\right)-\pi_p\left(h,l\right)+w(pm|f)-w(pm|s)\right]} \\ &> 0.\end{aligned} \tag{11}$$

This shows, if condition (11) holds, then the *assistant* optimally chooses *high* effort following the *selection tournament* and *high* effort by the *project manager*.

To summarize again, given that the principal uses the *selection tournament* to take his decision the *assistant* chooses *concentration* and:

1. *low* effort if the *project manager* has chosen *low* effort.
2. *high* effort if the *project manager* has chosen *high* effort and condition (11) holds.
3. *low* effort if the *project manager* has chosen *high* effort and condition (11) does not hold.

This brings us to the optimal behavior of the *project manager*. Consider first the *project manager*'s optimal behavior following the principal's choices to take the decision *behind closed doors*. From the above we know that the *project manager* and the principal know that the *assistant* always chooses *low* effort under *bcd*. Given this, the *project manager*'s perceived kindness of the principal's *procedural choice bcd*is:

$$\lambda_{pmppm} = w(pm|f) - \frac{1}{2}\left(w(pm|f) + \frac{1}{2}\left[(w(a) - v) + (w(pm|s) - v)\right]\right). \tag{12}$$

As $w(pm|f) = w(a)$, equation (12) reduces to:

$$\lambda_{pmppm} = w(a) - \frac{1}{2}\left(w(a) + \frac{1}{2}\left[(w(a) - v) + (w(pm|s) - v)\right]\right) < 0, \tag{13}$$

which is identical to equation (2). Hence, as the *assistant*'s optimal behavior is known to the *project manager* and the *project manager* also knows that the principal knows, the *project manager*' perceived kindness of the principal is identical to the assistant's perception following *bcd*. The same holds true for the *project manager*'s kindness. Given the optimal behavior of the *assistant*, the *project manager*'s kindness towards the principal reduces to:

$$\kappa_{pmp}(l) = (\pi_p(l,l) - w) - \frac{1}{2}\left((\pi_p(h,l) - w) + (\pi_p(l,l) - w)\right) = 0, \tag{14}$$

if he chooses *low* effort or

$$\kappa_{pmp}(h) = (\pi_p(h,l) - w) - \frac{1}{2}\left((\pi_p(h,l) - w) + (\pi_p(l,l) - w)\right) = 0, \tag{15}$$

if his effort choice is *high*. Concluding, as effort is costly also the optimal behavior of the *project manager* is *low* effort following the principal's procedural choice of *bcd*. What about the *selection tournament*? Remember, the assistant chooses *concentration* and l given that the pm chooses l and h if the pm chooses h and condition

(11) holds. Hence, the *project manager*'s perceived kindness following the *selection tournament* is:

$$\begin{aligned}\lambda_{pmppm} &= \frac{1}{2}\left[\left(w\left(a\right)-\upsilon\right)+\left(w(pm|s)-\upsilon\right)\right] \\ &- \frac{1}{2}\left(w(a)+\frac{1}{2}\left[\left(w\left(a\right)-\upsilon\right)+\left(w(pm|s)-\upsilon\right)\right]\right)>0. \end{aligned} \tag{16}$$

As can easily be seen, the *project manager* and the *assistant* feel equally treated. Hence, the perceptions about the principals kindness are identical (equations (16) and (7)). The *project manager* kindness towards the principal, on the other hand, following is:

$$\kappa_{pmp}(l)=\left(\pi_p\left(l,l\right)-w\right)-\frac{1}{2}\left(\left(\pi_p\left(h,h\right)-w\right)+\left(\pi_p\left(l,l\right)-w\right)\right)<0, \tag{17}$$

if he chooses *low* effort and

$$\kappa_{pmp}(h)=\left(\pi_p\left(h,h\right)-w\right)-\frac{1}{2}\left(\left(\pi_p\left(h,h\right)-w\right)+\left(\pi_p\left(l,l\right)-w\right)\right)>0, \tag{18}$$

if his effort choice is *high*. From this follows that the *project manager* chooses *concentration* and *high* effort following the *selection tournament*, if:

$$u_{pm}(h)\geq u_{pm}(l), \tag{19}$$

which can also be written as

$$\left(w(pm|s)-\upsilon\right)+Y_{pmp}\left(\kappa_{pmp}(h)\lambda_{pmppm}\right)\geq w(pm|f)+Y_{pmp}\left(\kappa_{pmp}(l)\lambda_{pmppm}\right). \tag{20}$$

This reduces to:

$$Y_{pmp}\geq\frac{\left(w(pm|f)-w(pm|s)\right)+\upsilon}{\lambda_{pmppm}\left(\kappa_{pmp}(h)-\kappa_{pmp}(l)\right)}. \tag{21}$$

Plugging in for the perceived kindness, λ_{pmppm}, and kindness, κ_{pmp} gives:

$$Y_{pmp} \;\geq\; \frac{\left(w(pm|f)-w(pm|s)\right)+\upsilon}{\frac{1}{2}\left[\frac{1}{2}\left[w\left(pm|s\right)-w(a)\right]-\upsilon\right]\left[\pi_p\left(h,h\right)-\pi_p\left(h,l\right)+w(pm|f)-w(pm|s)\right]}.$$

One can easily see that:

$$Y_{ap}\geq Y_{pmp}, \tag{22}$$

as $\left(w(pm|f)-w(pm|s)\right)<0$. Hence, the *project manager* optimally chooses *concentration* and *high* effort following the *selection tournament* already at lower levels of sensitivity to reciprocity compared to the *assistant*. This is due to the fact that he gets a financial reward for bringing *high* effort compared to the *assistant* who only supports him within the realm of his normal work and gets $w(a)$ independent of the success or failure of the project. Summarizing:

1. if the principal chooses to take the decision *behind closed doors*, both players optimally choose *low* effort in line with their beliefs and, in addition,

2. if conditions (11) and (22) hold, both choose *concentration* and *high* effort following the *selection tournament*.

Assume that both conditions (11) and (22) hold. Given this it is easy to see that the profit maximizing *principal* always chooses the *selection tournament* to take his decision, as this gives him a profit of $\pi_p(h, h)$. This concludes the proof. ■