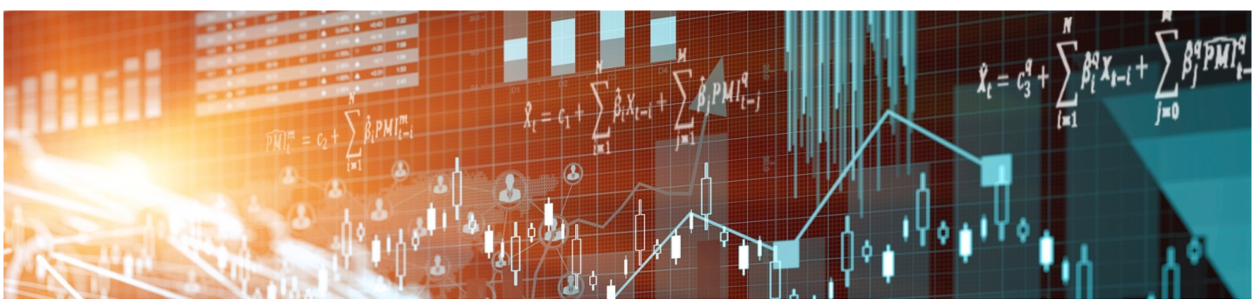
BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel 2019-50

Monetary Payoff and Utility Function in Adaptive Learning Models



by Erhao Xie

Bank of Canada Staff Working Paper 2019-50

December 2019

Monetary Payoff and Utility Function in Adaptive Learning Models

by

Erhao Xie

Funds Management and Banking Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
erhaoxie@bankofcanada.ca

ISSN 1701-9397

Acknowledgements

I thank Colin Camerer, Shota Ichihashi, Francisco Rivadeneyra, Ryan Webb, Yoram Halevy and seminar participants at the Bank of Canada, the University of Toronto and the Canadian Economics Association for helpful discussions. The views in this paper do not represent the views of the Bank of Canada.

Abstract

This paper focuses on econometric issues, especially the common assumption that monetary payoff is subjects' actual utility, in estimating subjects' learning behaviors using experimental data. I propose a generalized adaptive learning model that nests commonly used learning rules. First, I show that for a wide range of model parameters, adding a constant to the utility function alters players' learning dynamics. Such a range includes commonly used learning models, such as experience-weighted attraction (EWA), payoff assessment and impulse-matching learning. This result implies that the usual treatment of monetary reward as the actual utility is potentially misspecified, in addition to the common concern of risk preference. To deal with such an issue, the econometric model specifies a player's utility as an unknown function of monetary payoff. It is estimated jointly with the generalized learning model. I show that they are jointly identified under weak conditions. Using the experimental dataset by Selten and Chmura (2008) and Chmura et al. (2012), I reject the null hypothesis that monetary payoffs are utility. Incorrectly imposing such a restriction substantially biases the learning parameters, especially the weight on forgone utility. In addition, when a generalized model is considered, subjects are found to depreciate the unchosen action's experience more than the chosen one. Consequently, they are more responsive to the unchosen action's recent utility, rather than that of the chosen one. This feature is absent in commonly used learning models, such as EWA.

Bank topics: Econometric and statistical methods; Economic models
JEL codes: C57, C72, C92

1 Introduction

Players' learning behaviors have an important implication for their actual choices in games, especially when identical or similar games are played repeatedly. Specifically, learning relates to at least three core questions: Will players' behaviors converge to a Nash equilibrium or other solution concepts? If they converge, which equilibrium will they reach, and what is the speed of convergence? Because of the importance of this, economists have proposed a variety of models for understanding players' learning behaviors.[1] The performance of these models is usually evaluated by experimental data. In this literature, researchers commonly assume that the monetary payoffs assigned by experimentalists are subjects' actual utilities. This paper both argues that such an assumption is misspecified and provides a solution to the misspecification.

Assuming monetary payoff is the actual utility places strong restrictions on subjects' preferences. First, subjects must be risk neutral if this assumption is true. However, Goeree et al. (2003) and Harrison and Rütstrom (2008) provide evidence of risk preference at the payment scale in common experiments. Second, besides the concern regarding risk preference, this paper provides another reason why monetary payoff cannot be regarded as the actual utility. Many learning models, such as Mookherjee and Sopher (1994, 1997), Erev and Roth (1998), Camerer and Ho (1999) and Sarin and Vahid (1999, 2001), distinguish between *actual utility* and *forgone utility*. Specifically, players will discount the forgone utility of the strategy they have not chosen before.[2] I show that such a feature implies that adding a constant to the utility function will change subjects' learning path and, potentially, the converging outcome. In more detail, let $u(m)$ denote the utility function of monetary payoff m. Even when players are risk neutral, restricting $u(m) = m$ can still be misspecified. This is because another risk-neutral utility function $u(m) = m + c$ for some $c \neq 0$ can generate a different dynamic of choices. Consequently, researchers need to impose a correct value of c even though players can reasonably be assumed to be risk neutral.

At first glance, that a learning path is variant to adding a constant to the utility function seems to be a surprising result. Intuitively, if players value actual and forgone utilities differently, their learning

[1] A partial list of important contributions includes belief-based learning by Brown (1951), Fudenberg and Levine (1995, 1998) and Cheung and Friedman (1997); reinforcement learning by Erev and Roth (1998) and Mookherjee and Sopher (1994, 1997); experience-weighted attraction model by Camerer and Ho (1999), Camerer et al. (2002) and Ho et al. (2007); payoff assessment learning by Sarin and Vahid (1999), Sarin and Vahid (2001) and Cominetti et al. (2010); and impulse-matching learning by Chmura et al. (2012, 2014).

[2] The discount can be due to either incomplete information about the payoff or inattention.

paths depend not only on the history of each player's actions but also on the realized past utility. This is similar to the fallacy of sunk cost. Specifically, players make their current decisions based on their realized past utility even though it has no impact on their current expected utility. This feature implies that adding a constant to the utility will induce a different learning path. Finally, such a result has its own interest and can be used to test different versions of learning models. Specifically, the belief-based model does not distinguish between actual and forgone utility. Therefore, the learning dynamic depends only on the history of actions and is invariant to adding a constant to the utility function, in contrast to the reinforcement learning and payoff assessment models listed above. It consequently provides a testable restriction of belief-based learning against other models.

The constant c added to the utility function has a structural interpretation. Take the reinforcement learning model as an example; players will choose an action more/less frequently in the future if such an action is successful/unsuccessful in the current period. In this context, assuming monetary payoff as the utility restricts zero dollars to be the division between a successful and an unsuccessful strategy. In more detail, an action with a positive/negative monetary reward is coded as successful/unsuccessful. This restriction is implausible in many games.[3] The constant c is then interpreted as *how successful* does a player view the action with zero monetary reward. Such a constant is potentially unknown to researchers but could be estimated from players' choices.

Incorrectly assuming monetary payoff as the actual utility could generate considerable bias on the estimated learning parameters. Moreover, even though researchers know the learning parameters, we could still make incorrect counterfactual predictions. To address such an issue, this paper considers a general adaptive learning rule that nests most existing models as special cases. In addition, players' utility is specified as an unknown non-parametric function of the monetary payoff. Such a utility function is estimated together with the parameters of the general learning model. I show that under weak conditions, the utility function and learning parameters are jointly identified.

The experimental literature has developed several methods to relax the assumption of monetary payoff as utility. Roth and Malouf (1979) propose linearizing utility function by assigning the payoff as the probability of winning a fixed reward. This mechanism has been applied by Ochs (1995) and Fel-

[3]For instance, in a coordination game where both players can receive a positive monetary reward in some equilibria, an action that earns zero or a small positive amount of dollars is likely to be coded as an unsuccessful strategy.

tovich (2000), among others. Another common method is to elicit players' risk preference using a lottery choice. Such a method is used in Heinemann et al. (2009), among others. Even though these methods can adequately address risk preference, they are not suitable for estimation in the adaptive learning model. As described above, the knowledge of risk preference is not sufficient to consistently estimate learning parameters since adding a constant to the utility function could change the dynamics of learning behaviors.

Cabrales and Garcia-Fontes (2000) and Bracht and Ichimura (2001) study the identification and estimation of the *experience-weighted attraction* (EWA) model by Camerer and Ho (1999), but under the assumption that utility is the monetary payoff and therefore known by econometricians. This paper emphasizes the misspecification of such an assumption and specifies utility as an unknown function of money. This poses an additional identification burden on the model. Moreover, I consider a generalized model that nests EWA as a special case.

The identification results developed in this paper seem to conflict with the common wisdom that learning models are usually imprecisely estimated. Specifically, Salmon (2001) conducts a Monte Carlo study and shows that experimental data are insufficient to distinguish between different learning models. However, his study assumes a sample size of 40 subjects, which was a common experimental size at that time. Recently, many studies, such as Feri et al. (2010), Selten and Chmura (2008), Chmura et al. (2014), have conducted large-scale experiments with around 1,000 subjects. Using a Monte Carlo experiment, I show that researchers can reliably estimate both learning parameters and utility function with such an experimental scale. Moreover, incorrectly assuming monetary payoff as the utility generates considerable bias in both estimated learning parameters and counterfactual predictions.

In summary, this paper argues that researchers should be cautious when assuming monetary payoff as the actual utility. Such an assumption can be misspecified and leads to serious bias. When researchers have a large-scale experiment, a general learning model that nests commonly used learning procedures can be reliably estimated together with subjects' utility functions. Finally, using actual experiments by Selten and Chmura (2008) and Chmura et al. (2012), I reject the null hypothesis that the monetary payoff equals the actual utility. Players are risk averse and code a zero-dollar reward as unsuccessful.

The rest of this paper is organized as follows. Section 2 describes the model and Section 3 shows the identification results. I present the Monte Carlo experiment in Section 4 and the empirical results of an actual experiment in Section 5. Section 6 concludes.

2 Model

This section first presents the general adaptive learning model considered in this paper. This model nests commonly used learning procedures, such as the EWA model proposed by Camerer and Ho (1999), the payoff assessment learning studied by Sarin and Vahid (1999, 2001) and Cominetti et al. (2010), and the impulse-matching learning analyzed by Chmura et al. (2012), as special cases.

Consider an n-player normal-form game. Each player $i \in \{1,2,\cdots,n\}$ has J_i actions, and $\mathcal{S}_i = \{s_i^1, s_i^2, \cdots, s_i^{J_i}\}$ denotes the action set. An action/strategy is indexed by j. Consequently, the Cartesian product $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2 \cdots \times \mathcal{S}_n$ represents the set of action profiles in this game. Denote $\mathbf{s} \in \mathcal{S}$ as an arbitrary action profile; player i receives a payoff/utility $\pi_i(\mathbf{s})$ when $\mathbf{s}$ is the realized outcome of the game. This normal-form game is repeated for T periods. At period t, let $s_i(t)$ and $\mathbf{s}(t)$ denote the actual choice of player i and the realized outcome of the game, respectively. Then $\pi_i[\mathbf{s}(t)]$ represents player i's actual utility received in period t.

Researchers design an experiment that assigns each player's monetary reward for every action profile. Specifically, let $m_i(\mathbf{s})$ denote player i's monetary payoff when the realized outcome is $\mathbf{s}$. The existing literature commonly assumes the monetary payoff is the players' utility; for instance, $\pi_i(\mathbf{s}) = m_i(\mathbf{s})$. In contrast, this paper makes a distinction between these two terms. Even though $m_i(\cdot)$, referred to as the monetary payoff, is perfectly observed by researchers, term $\pi(\cdot)$, referred to as the utility, is unobserved. I assume only that the utility is strictly increasing in monetary payoff; it is summarized by Assumption 1.

Assumption 1. $\pi_i(\mathbf{s}) = u[m_i(\mathbf{s})]$ *where* $u(\cdot)$ *is a non-parametric, strictly increasing function.*

Note that I assume each player has the same utility function, for simplicity; the model and identification results are easily generalized to heterogeneous $u(\cdot)$ across players.

Each strategy has a numerical value, referred to as attraction, that determines the probability of choosing that action. An adaptive learning rule describes the evolution of these attractions and how the choice probability of each action depends on attractions. Specifically, let $A_{i,t}^j$ denote player i's attraction of strategy s_i^j at period t. It evolves according to

$$A_{i,t+1}^j = \frac{\phi[t, \mathbb{1}(s_i(t) = s_i^j)]}{N(t)} A_{i,t}^j + \frac{f(s_i^j, \mathbf{s}_t)}{N(t)} \pi_i(s_i^j, \mathbf{s}_{-i}(t)), \tag{1}$$
$$= \frac{\phi[t, \mathbb{1}(s_i(t) = s_i^j)] A_{i,t}^j + f(s_i^j, \mathbf{s}_t) \pi_i(s_i^j, \mathbf{s}_{-i}(t))}{N(t)}.$$

As shown in the first line, the future attraction of action s_i^j can be seen as a weighted sum of its current attraction and utility given other players' choices $\mathbf{s}_t$. Both $\phi(\cdot)$ and $N(\cdot)$ are non-parametric functions; therefore, I put no restrictions on how these weights will evolve across time. In addition, $\mathbb{1}(\cdot)$ is the indicator function that equals 1 if the statement in parentheses is true. Therefore, the weight on current attraction $A_{i,t}^j$ is allowed to differ based on whether action s_i^j is actually chosen or not. Finally, $f(\cdot)$ is the weight on the current utility, up to normalization of $N(t)$. It takes the following expression:

$$f(s_i^j, \mathbf{s}_{-i}(t)) = \begin{cases} 1 \text{ if } s_i(t) = s_i^j, \\ \delta_0 \text{ if } s_i(t) \neq s_i^j \text{ and } \pi_i[s_i^j, \mathbf{s}_{-i}(t)] < \pi_i[s_i(t), \mathbf{s}_{-i}(t)], \\ \delta_1 \text{ if } s_i(t) \neq s_i^j \text{ and } \pi_i[s_i^j, \mathbf{s}_{-i}(t)] \geq \pi_i[s_i(t), \mathbf{s}_{-i}(t)]. \end{cases} \tag{2}$$

If s_i^j is chosen at period t, then $\pi_i[s_i^j, \mathbf{s}_{-i}(t)]$ represents the realized/actual utility received by player i and it is updated with full weight. On the other hand, when s_i^j is not chosen, $\pi_i[s_i^j, \mathbf{s}_{-i}(t)]$ represents player i's utility if she deviates to s_i^j, holding other players' actions constant at period t. It is referred to as the forgone utility. Due to a player's limited attention, regret and/or potentially incomplete information about the preference, $\pi_i[s_i^j, \mathbf{s}_{-i}(t)]$ will be updated with a multiplier δ. Specifically, $\delta < 1$ if players pay less attention to the unchosen strategy, and $\delta > 1$ if players regret not choosing s_i^j. If player i has incomplete information such that the forgone utility is unknown to her, δ would be zero. In addition, depending on whether the forgone utility is higher than the actual utility $\pi_i[s_i(t), \mathbf{s}_{-i}(t)]$, δ is allowed to take different values. This captures the possibility that players could pay more attention to the unchosen action that yields a higher utility than the current chosen strategy. For instance, Ho et al. (2007) assume $\delta_1 = 1$ and $\delta_0 = 0$.

In a similar vein to the EWA model by Camerer and Ho (1999), the second line of equation (1) provides another interpretation of the evolution. Specifically, $N(t)$ can be interpreted as the number of "observation-equivalent" stage games that a player has played. Term $\phi(\cdot)A_{i,t}^{j}$ then represents the cumulative attraction of action s_i^j at period t. The weight $f(\cdot)$ describes how the current utility of s_i^j is accumulated into future attractions. Consequently, $A_{i,t+1}^{j}$ can be interpreted as a weighted average of historical utilities for action s_i^j up to period t.

The attractions in each period determine players' choice probabilities of every action. Let $P_{i,t}^{j}$ denote player i's choice probability of action s_i^j at period t. As standard in the literature, I specify a logit formula:

$$P_{i,t}^{j} = \frac{\exp(\lambda A_{i,t}^{j})}{\sum_{k=1}^{J_i} \exp(\lambda A_{i,t}^{k})}. \tag{3}$$

As in McKelvey and Palfrey (1995), the logit formula can be seen as player i chooses an action that maximizes the following:

$$\max_{s_i \in \mathbb{S}_i} \tilde{A}_{i,t}^{s_i} = A_{i,t}^{s_i} + \varepsilon_{i,t}^{s_i},$$

where $\varepsilon_{i,t}^{s_i}$ is a random shock to attraction of s_i^j at period t. It follows the Gumbel distribution and is independent across time and actions. Moreover, λ measures the inverse of the standard deviation of $\varepsilon_{i,t}^{s_i}$. It is referred to as the sensitivity parameter. If player i best responds to attraction (i.e., chooses the strategy with the highest attraction), then $\lambda = \infty$. In contrast, $\lambda = 0$ implies that players are insensitive to attraction and randomly choose different actions with equal probabilities. Finally, following the discrete choice literature, the identification results are easily generalized when ε follows other distributions, for instance, the multinomial probit model.

Subsections 2.1 to 2.3 show that the adaptive learning model considered in this paper nests commonly used learning rules as special cases. For the sake of brevity, the learning rule specified by equations (1) to (3) is referred to as the ***Generalized Adaptive Learning*** (GAL) model.

2.1 Experience-weighted Attraction (EWA)

Camerer and Ho (1999) propose the EWA model. It is the benchmark in the literature and shown to fit experimental data successfully.[4] Compared with the GAL model described above, EWA learning imposes

[4]See Camerer and Ho (1999), Ho et al. (2007), and Chmura et al. (2012).

three additional restrictions. First, the evolution of weights $N(\cdot)$ and $\phi(\cdot)$ are parametric functions of time t. Second, the weight on current attraction of action s_i^j is independent of whether s_i^j is chosen or not. Third, players discount the forgone utility equally regardless of whether it is higher than the actual utility or not. These three restrictions are represented by the following equations:

$$\begin{aligned} N(t) &= \rho \cdot N(t-1) + 1, \\ \phi[t, \mathbb{1}(s_i(t) = s_i^j)] &= \phi \cdot N(t), \\ \delta_1 = \delta_0 &= \delta, \end{aligned} \tag{4}$$

where ρ, ϕ and δ are unknown parameters. With these three restrictions, the GAL model becomes the EWA as follows:

$$N(t) = \rho \cdot N(t-1) + 1,$$
$$A_{i,t+1}^j = \frac{\phi N(t) A_{i,t}^j + [\delta + (1-\delta) \cdot \mathbb{1}(s_i(t) = s_i^j)] \cdot \pi_i[s_i^j, \mathbf{s}_{-i}(t)]}{N(t)}.$$

Camerer and Ho (1999) further show that the EWA model nests the reinforcement learning of Mookherjee and Sopher (1994, 1997) and weighted fictitious play in Brown (1951), Fudenberg and Levine (1995, 1998) and Cheung and Friedman (1997). Consequently, these models are also nested in the GAL model.

2.2 Payoff Assessment Learning

Sarin and Vahid (1999, 2001) study a payoff assessment learning model. When s_i^j is chosen, its future attraction is a weighted average of its current attraction and actual utility. In contrast, when s_i^j is not chosen, player i does not update its attraction at all and the same value of attraction transmits to the next period. Equivalently, the weight on the forgone utility is zero, and the weight on the current attraction $A_{i,t}^j$ depends on whether s_i^j is chosen or not. In addition, Cominetti et al. (2010) study a general version of payoff assessment learning that allows the weights to evolve across time.

Suppose we impose the following restrictions on the GAL model:

$$\phi(t,0) = N(t),$$
$$\frac{\phi(t,1)}{N(t)} + \frac{1}{N(t)} = 1, \tag{5}$$
$$\delta_1 = \delta_0 = 0.$$

The evolution of attraction turns to

$$A_{i,t+1}^{j} = \begin{cases} (1 - \frac{1}{N(t)})A_{i,t}^{j} + \frac{1}{N(t)}\pi_i[s_i^j, \mathbf{s}_{-i}(t)], \text{ if } s_i(t) = s_i^j, \\ A_{i,t}^{j}, \text{ otherwise.} \end{cases}$$

This is the evolution considered in Cominetti et al. (2010). It also nests Sarin and Vahid (1999, 2001). Consequently, payoff assessment learning can be seen as a special case of the GAL model.

2.3 Impulse-matching Learning

Chmura et al. (2012) consider an impulse-matching learning in the context of a two-player binary-choice game. They argue that players would value gains differently than losses, according to the prospect theory proposed by Kahneman and Tversky (1979). Specifically, let $m_i^* = \max_{s_i \in \mathcal{S}_i}\{\min_{\mathbf{s}_{-i}} m_i(s_i, \mathbf{s}_{-i})\}$. It represents the maximal monetary payoff that player i can obtain for sure in each period. Chmura et al. (2012) argue that m_i^* forms a natural aspiration level. Any monetary payoff lower than it is perceived as a loss, while a higher one is regarded as a gain. In line with Kahneman and Tversky (1979), they assume losses are counted double in comparison with gains. Specifically, for each action profile $\mathbf{s}$, they assume the following utility function:

$$\pi_i(\mathbf{s}) = \begin{cases} \frac{m_i(\mathbf{s}) + m_i^*}{2}, \text{ if } m_i(\mathbf{s}) \geq m_i^*, \\ m_i(\mathbf{s}), \text{ if } m_i(\mathbf{s}) < m_i^*. \end{cases} \tag{6}$$

The utility function is strictly increasing in monetary payoff and therefore is nested in Assumption 1.[5]

[5] Given equation (6), players have heterogeneous utility functions as they have different aspiration levels m_i^*. This seems to

Consider the following restrictions imposed in the GAL model:

$$N(t) = \phi[t, \mathbb{1}(s_i(t) = s_i^j)] = 1,$$
$$\delta_1 = \delta_0 = 1. \tag{7}$$

The evolution of attraction then becomes

$$A_{i,t+1}^j = A_{i,t}^j + \pi_i[s_i^j, \mathbf{s}_{-i}(t)].$$

Such an evolution seems to be different from the following rule specified in Chmura et al. (2012):

$$A_{i,t+1}^j = A_{i,t}^j + \max\left\{0, \pi_i[s_i^j, \mathbf{s}_{-i}(t)] - \pi_i[s_i^k, \mathbf{s}_{-i}(t)]\right\}.$$

However, under the logit formula given by equation (3), only the difference in the two strategies' attractions determines the choice probability. Moreover, it is easy to see that both evolutions yield the same difference as the following:

$$A_{i,t+1}^j - A_{i,t+1}^k = A_{i,t}^j - A_{i,t}^k + \pi_i[s_i^j, \mathbf{s}_{-i}(t)] - \pi_i[s_i^k, \mathbf{s}_{-i}(t)].$$

Therefore, when researchers impose logit choice probability, the impulse-matching learning by Chmura et al. (2012) is nested in the GAL model.[6]

2.4 Parametrized Generalized Adaptive Learning

The next section shows that when researchers have a perfect dataset such that the number of experimental subjects approaches infinity, utility function and transitional functions can be identified. However, in practice, experimentalists recruit only a finite and usually small number of subjects, due to time and financial constraints. In this case, researchers have to make a compromise and consider a parsimonious model to achieve precise estimation. This subsection proposes parameterizing the transitional function to

be contradicted by Assumption 1, in which the preference is assumed to be homogeneous. However, as described above, the model and identification results in this paper are easily generalized when the utility function is player specific.

[6] Chmura et al. (2012) apply a power probability with power equal to 1.

ease the estimation burden. Similar to the EWA model, I consider the following restrictions:

$$
\begin{aligned}
N(t) &= \rho N(t-1)+1, \\
\phi(t,0) &= \phi_0 \cdot N(t-1), \\
\phi(t,1) &= \phi_1 \cdot N(t-1).
\end{aligned}
\tag{8}
$$

The evolution of $N(t)$ is identical to the EWA model. In contrast, $\phi N(t-1)$ represents the weight on the current attraction. The EWA model restricts ϕ to be constant across all actions, while I allow ϕ to be different for the chosen and unchosen strategies. Comparing equation (8) with equations (4), (5) and (7), it is easy to see that the EWA, payoff assessment learning and impulse-matching models are also nested in this parametrized GAL model.

3 Identification

This section first presents conditions for the data-generating process. Subsection 3.2 derives some empirical properties of the GAL model; they lead to some normalizations for identification. One important property is that adding a constant to the utility can alter players' learning path. Subsection 3.3 presents the identification results, and Subsection 3.4 discusses individual heterogeneity. All proofs are in the Appendix.

3.1 Data-Generating Process

Suppose researchers recruit in total nG subjects and divide them into G groups. Each subject is randomly assigned one role of n players. The role is fixed, and each subject plays a sequence of normal-form games described in Section 2 for T periods. Groups can be either fixed or randomly re-paired across time. Importantly, each subject learns according to the GAL model, as specified by equations (1) to (3). They are homogeneous so that each subject has the same utility function and learning rule. Specifically, $N(\cdot)$, $\phi(\cdot)$ and $u(\cdot)$ are identical across subjects. Subsection 3.4 discusses the identification results when individual heterogeneity is allowed in the econometric model.

This paper focuses on players' learning behaviors. Therefore, I assume each action will be chosen with a positive probability in each period.[7] Suppose a player sticks to one particular action; it can be interpreted as this player stops learning. Such a scenario is not the focus of this paper. The asymptotics come from the number of subjects approaching infinity while T is fixed; equivalently, $G \to \infty$. Denote $P_{i,t}^{j}[\mathbf{s}(1), \cdots, \mathbf{s}(t-1)]$ as player i's choice probability of s_i^j given an arbitrary history $[\mathbf{s}(1), \cdots, \mathbf{s}(t-1)]$. Such a choice probability can be consistently estimated. This is because any history can be observed infinitely many times when $G \to \infty$. Consequently, I assume $P_{i,t}^{j}[\mathbf{s}(1), \cdots, \mathbf{s}(t-1)]$ is known to researchers. The objective is to use this type of experimental data to identify unknown functions of the GAL model and utility, for instance, transitional function $N(\cdot)$ and $\phi(\cdot)$, weighting function $f(s_i^j, \mathbf{s}(t))$, utility function $u(\cdot)$ and initial attraction $A_{i,1}^{j}$.

3.2 Empirical Properties and Normalizations

It is straightforward to show that any transformations in Lemma 1 will leave $\lambda A_{i,t}^{j}$ unchanged for any t. According to equation (3), a player's choice probability is invariant to any of these transformations. This result calls the normalizations in Assumption 2.

Lemma 1. *Player i' choice probability $P_{i,t}^{j}[\mathbf{s}(1), \cdots, \mathbf{s}(t-1)]$ is invariant to any of the following transformations:*

(a) Multiply a constant c to initial attraction $A_{i,1}^{j}$ and utility function $u(\cdot)$, then multiply the sensitivity parameter λ by a constant $\frac{1}{c}$.

(b) Multiply $u(\cdot)$, $N(\cdot)$, and $\phi(\cdot)$ by the same constant c.

Assumption 2. *Lemma 1 leads to the following normalizations: $\lambda = 1$ and $N(1) = 1$.*

As argued by McKelvey and Palfrey (1995), λ measures the players' sensitivity to the expected utility or attraction. However, as stated in Lemma 1 (a), it is indistinguishable from the scale of initial attraction and utility function. For example, suppose a subject's action is extremely sensitive to the payoff of \$1. Such a behavior can be equivalently explained by either the subject having a high sensitivity parameter λ or the utility of \$1 being extremely high. Previous literature identifies λ because it assumes utility is

[7] Such an assumption can hold under weak conditions. For instance, it is true if λ and attraction are finite numbers.

money; for instance, the utility of \$1 is 1. This paper non-parametrically specifies the utility function and therefore normalizes λ to unity. Finally, this paper presents the model that focuses on a single normal-form game. Suppose researchers design an experiment with multiple types of normal-form games where each type of game has a different monetary payoff matrix. Under the assumption that utility function is constant across all games, the identification results studied in this paper identify how the λ varies across different types of games. Consequently, researchers can test the hypothesis, for instance, whether players are more sensitive to attractions in coordination games than in matching pennies games. Moreover, normalizing $\lambda = 1$ in only one game is enough for point identification.

Camerer and Ho (1999) estimate initial experience $N(1)$; they achieve the identification because they parametrize the transition function–for instance, $N(t) = \rho N(t-1) + 1$. In contrast, this paper non-parametrically specifies such a transition and normalizes $N(1) = 1$. More importantly, even with parametrization, $N(1)$ is usually poorly estimated, and assuming $N(1)$ is common in the literature, for instance, Ho et al. (2007).

Proposition 1. *Under Assumptions 1 and 2, player i's choice probability* $P^j_{i,t+1}[\mathbf{s}(1),\cdots,\mathbf{s}(t)]$ *is invariant to adding a constant to the utility function* $u(\cdot)$ *if and only if the following two conditions hold:*

(a) $\delta_1 = \delta_0 = 1$*; i.e.* $f(s^j_i, \mathbf{s}) = 1$ *for each* $s^j_i \in \mathcal{S}_i$ *and* $\mathbf{s} \in \mathcal{S}$.

(b) $\delta(t,1) = \delta(t,0)\ \forall t > 1$.

Under standard preference theory, adding a constant to the utility of each action has no impact on the optimal decision; consequently, the choice probability is invariant. As shown in Proposition 1, such a property is preserved only in some special cases of the GAL model. Specifically, conditions (a) and (b) imply that players put the same weight on chosen and unchosen strategies when they update their attractions; therefore, adding a constant to the utility has no effect on the evaluation of each strategy. However, if either condition fails, chosen and unchosen strategies will follow different updating rules. For instance, consider that condition (a) fails and $\delta < 1$. Suppose a positive constant c is added to the utility; then the chosen strategy receives the full weight of the actual utility and its future attraction will increase by c. In contrast, the unchosen strategy receives less than the full weight of the forgone utility, and its future attraction increases by less than c. As a result, the unchosen strategy will be chosen less frequently in the future, compared with the scenario in which c is not added to the utility. As shown in

Section 2, a broad class of learning models, such as reinforcement and payoff assessment, assume unequal weights on the chosen and unchosen actions; therefore, players' learning paths in all these models will be different if a constant is added to the utility. As a comparison, the belief-based learning model restricts an identical updating rule of attractions for all strategies; therefore, the invariance property of choice probability holds. By the above comparison, Proposition 1 also provides a testable implication of belief-based learning against other types of models.

3.3 Identification Results

This subsection presents the identification results. Proposition 2 states that transitional functions $N(\cdot)$, $\phi(\cdot)$ and initial attractions are identified under weak conditions.

Proposition 2. *Under Assumption 2 and supposing $T \geq 3$, we have the following:*

(a) $N(t)$, $\phi(t,0)$ and $\phi(t,1)$ are point identified for each t.

(b) Initial attraction $A_{i,1}^{j}$ is identified for each player i and strategy s_i^j.

(c) Term $f(s_i^j,\mathbf{s})\pi_i(s_i^j,\mathbf{s}_{-i}) - f(s_i^k,\mathbf{s})\pi_i(s_i^k,\mathbf{s}_{-i})$ is identified for any pair of s_i^j, s_i^k and any $\mathbf{s} = (s_i,\mathbf{s}_{-i})$.

The identification of transitional functions only requires the normalizations in Assumption 2, and the stage game is repeated more than twice. Note that it does not rely on Assumption 1. Therefore, $N(\cdot)$ and $\phi(\cdot)$ are identified even when a player's utility function may depend on other players' monetary rewards. It consequently allows the social preference, as studied by Güth et al. (1982), Kahneman et al. (1986) and Fehr and Schmidt (1999), etc.

Term $f(s_i^j,\mathbf{s})\pi_i(s_i^j,\mathbf{s}_{-i}) - f(s_i^k,\mathbf{s})\pi_i(s_i^k,\mathbf{s}_{-i})$ represents the weighted difference of any two actions' utilities. The remainder of this subsection shows how to separately identify the utility function and the weight $f(\cdot)$ under Assumption 1.

Suppose researchers design a variety of treatments in which each treatment constitutes a different monetary payoff matrix. Moreover, every monetary reward can be received by a subject with positive probability as the number of treatments approaches infinity. Such a data-generating process directly identifies utility function and weight given the information on weighted difference of utility. However, due to time and financial constraints, experimentalists commonly study a small number of treatments, and the number of possible monetary payoffs is finite. This paper presents a sufficient condition to identify the utility function under such a data-generating process. It therefore provides guidance for researchers to design an experiment.

Proposition 3. *Under Assumptions 1, 2 and with $T \geq 3$, suppose further that $\phi(t,0) \neq \phi(t,1)$ for some $1 < t < T$; then the utility function $\pi_i(\mathbf{s}) = u[m_i(\mathbf{s})]$ is identified for any $\mathbf{s} \in \mathcal{S}$. Moreover, the weight function $f(s_i, \mathbf{s})$ is identified for any s_i and $\mathbf{s}$.*

Besides Assumptions 1 and 2, Proposition 3 requires only that $\phi(t,0) \neq \phi(t,1)$ for some t. Such a condition is satisfied in payoff assessment learning. In this model, the chosen action receives a different weight on its attraction than the unchosen strategy. Moreover, given Proposition 2, $\phi(\cdot)$ is identified; therefore, the condition that $\phi(t,0) \neq \phi(t,1)$ is testable. When the above condition holds, the utility function is identified for any monetary payoff matrix.

There exists a broad class of models that restrict $\phi(t,0) = \phi(t,1)$–for instance, the EWA learning. Under such a scenario, additional restrictions have to be imposed on the monetary payoff matrix to achieve the identification. To describe these restrictions, some burdensome notations have to be introduced. The description and restrictions summarized by Assumption 3 are included in the Appendix.

Proposition 4. *Suppose that $\phi(t,0) = \phi(t,1)$ under Assumptions 1 to 3 and $T \geq 3$; we then have the following:*

(a) The weight function $f(s_i, \mathbf{s})$ is identified for any s_i and $\mathbf{s}$.

(b) If $f(s_i, \mathbf{s}) \neq 1$ for some $(s_i, \mathbf{s})$, then utility function $u[m_i(\mathbf{s})]$ is identified for any $\mathbf{s}$.

(c) If $f(s_i, \mathbf{s}) = 1$ for all $(s_i, \mathbf{s})$, then the difference of utility $u[m_i(s_i^j, \mathbf{s}_{-i})] - u[m_i(s_i^k, \mathbf{s}_{-i})]$ is identified for any pair of s_i^j, s_i^k and any choice of other players $\mathbf{s}_{-i}$.

If weight function $f(\cdot)$ equals 1 for any action profile, under restriction such that $\phi(t,1) = \phi(t,0)$, players' learning path is invariant to adding a constant to the utility function as shown in Proposition 1. In this case, a player's choice probability depends on the difference of two actions' utilities, rather than the absolute value. Therefore, such a difference is the only utility primitive that can be identified from the data. If $f(\cdot) \neq 1$ for some action profile, the utility of every possible monetary reward is identified.

3.4 Individual Heterogeneity

The previous subsection assumes homogeneous subjects with identical learning rules and utility functions. Consequently, the identification results directly imply that model primitives can be consistently estimated when the number of subjects approaches infinity. However, individuals can have heterogeneous preferences and learning parameters. The existence of individual heterogeneity, as shown by Wilcox (2006), produces a bias that tends to favor the reinforcement learning relative to the belief-based model.

A natural way to deal with individual heterogeneity is to consider the other extreme of the data-generating process, as studied in Cabrales and Garcia-Fontes (2000). Suppose researchers impose no restrictions on the distribution of individual heterogeneity; is it possible to consistently estimate learning rules and utility functions at the individual level if a subject makes infinite choices (i.e., $T \to \infty$)? The answer is obviously "no" in the non-parametric version of the GAL model. As the learning rule imposes no restrictions on the weights across time, the model has more unknowns than observations. In contrast, the parametrized GAL model restricts the evolution of weights. With an additional restriction such that the action profile chosen at period t has zero impact on period $t+h$'s choice probability when $h \to \infty$ (assumption of the independence of remote history), Cabrales and Garcia-Fontes (2000) show that the model primitives are consistently estimated, provided they are identified.[8] Since Propositions 2 to 4 establish the identification of primitives, the parametrized GAL model can allow individual heterogeneity.

There exist some practical issues. First, the number of periods T is typically small, and parameters are imprecisely estimated at the individual level. Second, the independence of remote history assumption restricts the converging outcome to be independent of initial choices; it rules out many interesting scenarios, such as the coordination game. In practice, researchers have to make a compromise by imposing additional restrictions on the distribution of individual heterogeneity. For example, Camerer and Ho (1999), Cabrales and Garcia-Fontes (2000) and Wilcox (2006) estimate a mixture model.

[8]Initial experience and initial attraction cannot be consistently estimated because remote future choices are independent of them.

4 Monte Carlo Experiment

This section presents a Monte Carlo experiment. It aims to illustrate the consequences of misspecifying the utility function. Specifically, the misspecification causes considerable bias in both estimation and counterfactual predictions. It also shows that, under the parametrization in Subsection 2.4, the GAL model and utility function can be reliably estimated in a dataset with 1,000 subjects. This is the sample size in many recent large-scale experiments, such as Feri et al. (2010), Selten and Chmura (2008), and Chmura et al. (2012, 2014). Throughout this subsection, I consider the parametrization given by equation (8).

This subsection focuses on a two-player binary choices coordination game shown in Table 1. The top panel shows the monetary payoff matrix. In an experiment, such a matrix is designed and observed by researchers. Specifically, each player has two choices. One is safe action X, which yields \$4 regardless of the other player's behavior. Y is a risky choice that generates \$9 if the other player also chooses Y, but only yields \$1 when the other selects the safe choice. This type of coordination game has received lots of attention in both the learning and the experimental literature, for instance, Van Huyck et al. (1990), Heinemann et al. (2009), Feri et al. (2010). It has two pure strategy Nash equilibria, an efficient equilibrium (Y,Y) and an inefficient one (X,X). A learning model plays an important role in this coordination game since it explains whether and which one of these equilibria can be reached in the long run.

I assume players are risk averse with a utility function $u(m) = \sqrt{m}$. Therefore, the middle panel in Table 1 represents the actual coordination game played by both players. Even though researchers perfectly observe each player's monetary payoff, they may not necessarily know the utility function. As a result, the middle panel and the actual game are potentially unobserved by researchers. The results in this paper can be used to reliably estimate the actual utility. The bottom panel, labeled "Incorrect Utility," assumes a utility function $u(m) = \sqrt{m} + 2$. Since players have the same risk preference, the middle and bottom panels represent the identical game and have the same set of Nash equilibria. However, the bottom one is "incorrect" in the sense that it misspecifies the location of the utility function (i.e., $u(0) = 2$ rather than $u(0) = 0$). In a wide range of parameter values in the GAL model, heterogeneous locations generate different choice dynamics and converging outcomes.

I study the case in which groups are fixed across time. The experiment considers four sets of pa-

Table 1: Monte Carlo Experiment: Coordination Game

Monetary Payoff

		Player 2	
		X	Y
Player 1	X	[4 , 4]	[4 , 1]
	Y	[1 , 4]	[9 , 9]

True Utility

		Player 2	
		X	Y
Player 1	X	[2 , 2]	[2 , 1]
	Y	[1 , 2]	[3 , 3]

Incorrect Utility

		Player 2	
		X	Y
Player 1	X	[4 , 4]	[4 , 3]
	Y	[3 , 4]	[5 , 5]

rameters, shown in Table 2. The first column presents the benchmark set of parameters. It assumes that players put more weight on both the current attraction and the actual utility of the chosen action. I choose this set of parameters since it ensures each group of subjects will converge to one of the pure strategy Nash equilibria.[9] The second column assumes that players have the same weight on the current attraction regardless of the action they actually choose. Moreover, actual and forgone utilities are updated equally. Given Proposition 1, the dynamics of players' choice probabilities are invariant to adding a constant to the utility function under such a scenario. Therefore, the middle and bottom games in Table 1 should generate the same learning path. Finally, the third column restricts the same weight on the actual and forgone utility while the last column assumes equal weight on each action's attraction. For each set of parameter values in Table 2 and each matrix in Table 1, I simulate the learning paths of 100,000 groups of two-player games. The number of periods is $T = 200$.

Figure 1 shows the fraction of efficient outcome (Y,Y) across periods.[10] The simulation can be seen

[9] In my simulation, more than 99% of groups converge to one of the equilibria.

[10] The simulation assumes that each player chooses Y with a low initial probability (i.e., 37%) such that the initial fraction of efficient equilibrium is 15%. I choose a low starting efficiency because it sheds light on differential learning paths and

Table 2: Monte Carlo Experiment: Values of Parameters

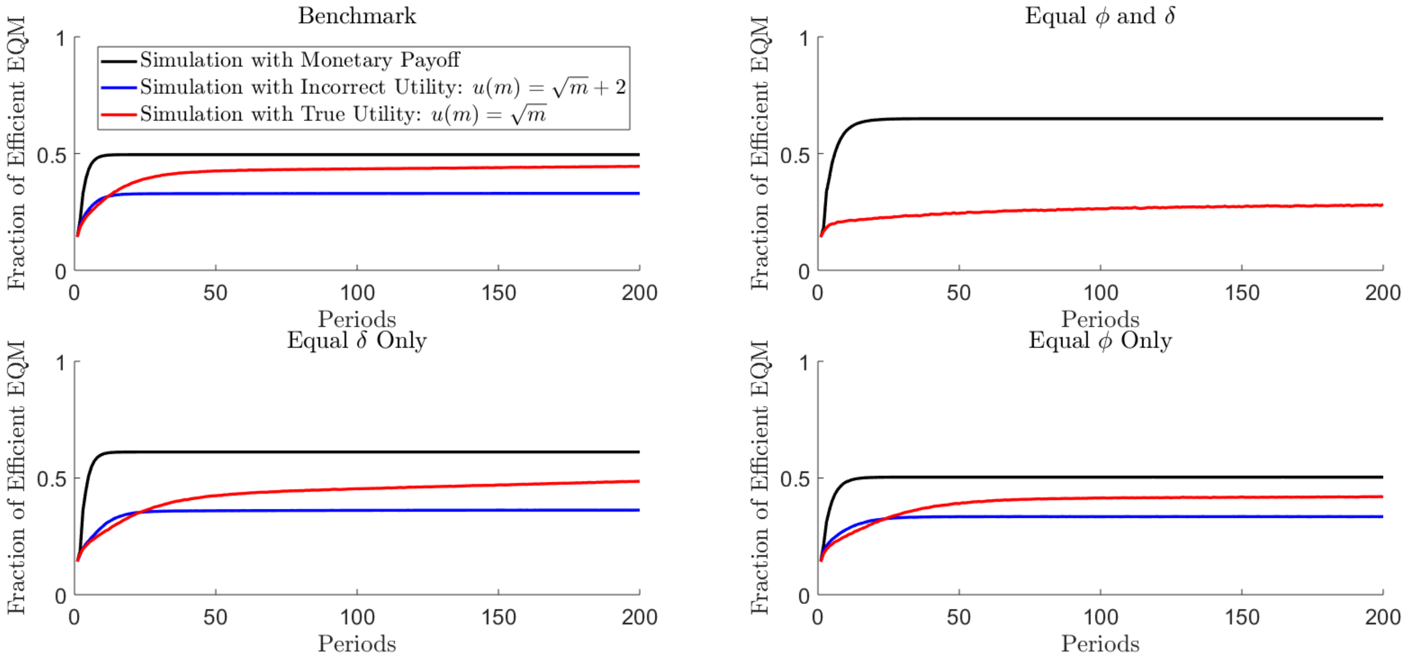
Parameter	Benchmark	Equal ϕ and δ	Equal δ	Equal ϕ
ρ	0.8	0.8	0.8	0.8
ϕ_0	0.7	0.9	0.7	0.9
ϕ_1	0.9	0.9	0.9	0.9
δ_0	0.5	1	1	0.8
δ_1	0.8	1	1	0.5

as an exercise of counterfactual predictions such that researchers know the values of learning parameters but do not know players' actual utility, as shown in the middle matrix of Table 1. It sheds light on the bias of misspecification of the utility function. The black, blue and red lines represent the choice paths under monetary payoff, incorrect utility and true utility, respectively. Therefore, the comparison between the black and red lines reflects the effect of risk aversion on the learning path, and the difference between the red and blue lines represents the effect of a constant on the utility function. As the benchmark shows, incorrectly assuming monetary payoff as utility predicts a quick learning behavior; the fraction of efficient outcome reaches 50% at around the 10th period and remains at that level thereafter. In contrast, if players are risk averse, they will learn at a slower speed, and less than half of the groups can reach an efficient equilibrium. Moreover, such a fraction decreases when a positive number is added to the utility function. If ϕ and δ are fixed across each action, as shown in the top right graph, adding a constant to the utility has no impact on the learning path. Therefore, the blue and red lines overlap. Finally, as shown in the last row, when either ϕ or δ is the same across actions, adding a positive constant to the utility generates a faster learning speed but a lower fraction of efficient equilibrium upon convergence.

The second Monte Carlo experiment studies whether researchers can reliably estimate the learning parameters and utility function. The experiment first simulates a dataset with $G = 500$ groups of subjects for $T = 50$ periods. This simulation uses the parameters shown in the benchmark column in Table 2 and true utility $u(m) = \sqrt{m}$. Since each group consists of two players, the sample size is 1,000. Given this simulated data, I estimate the learning parameters as well as the utility function. The experiment is repeated 100 times. In addition, the model is estimated by maximum likelihood estimation (MLE). Specifically, consider subject n. Given the learning parameters and utility function, her choice probabili-

converging outcomes for various sets of model parameters and payoff matrices. In contrast, players will remain at choice Y if they start from (Y,Y), for a wide range of parameters values and preferences.

Figure 1: Monte Carlo Experiment: Simulation Path



ties of two actions $P^X_{n,t}$ and $P^Y_{n,t}$ are well defined for each period t. Therefore, the model can be estimated by maximizing the following log-likelihood function:

$$LL = \sum_{n=1}^{1,000} \mathbb{1}(s_{n,t} = X)\log(P^X_{n,t}) + \mathbb{1}(s_{n,t} = Y)\log(P^Y_{n,t}). \tag{9}$$

Table 3 presents the average and standard deviation of estimates for 100 Monte Carlo samples. The first column assumes that researchers know the utility function $u(m) = \sqrt{m}$ and plug it into the log-likelihood. As true utility is usually unobserved by researchers, this specification is a hypothetical scenario and serves as a benchmark for comparison. The second column shows the estimates when researchers incorrectly assume monetary payoff as utility. First, there is moderate bias in transitional parameters ρ and ϕ. The bias is less than 10%. As the proof of Proposition 2 shows, the utility function does not appear into the identifying equations of ρ and ϕ. Therefore, the misspecification of $u(\cdot)$ has a relatively small impact on these two parameters. In contrast, there is substantial bias on the weight of the forgone utility. Specifically, δ_1 is biased down by 35% and δ_0 is biased up by more than three times. Moreover, it also leads to a qualitatively incorrect conclusion that $\delta_0 > 1 > \delta_1$. It implies that, compared with an unchosen action with high utility, a player regrets not choosing an action with low utility. Such

Table 3: Monte Carlo Experiment: Full Model Estimation Results

Parameters and True Values	True Utility	Monetary Payoff	Incorrect Utility	Parametric Utility	Non-parametric Utility
$\rho = 0.8$	0.7981 (0.0109)	0.8663 (0.0081)	0.8488 (0.0113)	0.7962 (0.0197)	0.7958 (0.0202)
$\phi_0 = 0.7$	0.6995 (0.0109)	0.7598 (0.0088)	0.7573 (0.0079)	0.6998 (0.0129)	0.7000 (0.0159)
$\phi_1 = 0.9$	0.8990 (0.0056)	0.8734 (0.0071)	0.8819 (0.0068)	0.8988 (0.0057)	0.8986 (0.0061)
$\delta_0 = 0.5$	0.5036 (0.0541)	1.7592 (0.0257)	0.7903 (0.0226)	0.5086 (0.1318)	0.5092 (0.1422)
$\delta_1 = 0.8$	0.7984 (0.0342)	0.5193 (0.0139)	0.9445 (0.0243)	0.7914 (0.0395)	0.7909 (0.0397)
$u(1) = 1$	*n.a.* *n.a.*	*n.a.* *n.a.*	*n.a.* *n.a.*	0.9871 (0.1324)	0.9902 (0.1693)
$u(4) = 2$	*n.a.* *n.a.*	*n.a.* *n.a.*	*n.a.* *n.a.*	2.0017 (0.2291)	2.0062 (0.2518)
$u(9) = 3$	*n.a.* *n.a.*	*n.a.* *n.a.*	*n.a.* *n.a.*	3.0273 (0.3923)	3.0343 (0.4254)

Notes: Numbers without parentheses are averages of estimates among 100 Monte Carlo samples. Numbers in parentheses are corresponding standard deviations.

an anomaly is the result of incorrectly assuming money as utility. As shown in the proof of Propositions 3 and 4, the identification of δ crucially depends on players' utility. Consequently, the misspecification of $u(\cdot)$ has a potentially serious impact on the weight of the forgone utility. The third column assumes an incorrect utility function $u(m) = \sqrt{m} + 2$. Specifically, researchers correct players' risk preference but impose an incorrect location (e.g., $u(0) = 2$ rather than $u(0) = 0$). Compared with the second column, correcting players' risk preference estimates a correct qualitative relationship with $\delta_0 < \delta_1 < 1$. However, a sizable bias still remains for both δ_0 and δ_1.

The fourth column assumes that researchers do not know the utility but know the functional form. Specifically, it assumes a utility function $u(m) = \alpha + m^{\beta}$, where α and β are unknown parameters that are estimated together with the learning model. As shown in the table, both utility and learning parameters are precisely estimated; their averages across Monte Carlo samples are close to true values. Compared with the first column, except for ρ and δ_0, all parameters' standard deviations increase by a negligible amount in the fourth column. Finally, the last column relaxes the assumption that researchers know the utility functional form. Instead, it specifies utility as a non-parametric function of monetary payoff. In more detail, $u(1)$, $u(4)$ and $u(9)$ are specified as three unknown parameters. As the last column shows, all parameters are reliably estimated. Compared with the fourth column, the standard deviations for all parameters increase by a small amount.

Since commonly used existing learning models can be seen as special cases of GAL considered in this model, my last Monte Carlo experiment studies how these additional restrictions increase the precision of estimated parameters. It then sheds light on the trade-off between a parsimonious and a generalized model. Specifically, I consider the restrictions such that $\phi_0 = \phi_1$ and $\delta_0 = \delta_1$. As shown in Subsection 2.1, these restrictions reduce the generalized model to the EWA model.

Table 4: Monte Carlo Experiment: Reduced Model (EWA) Estimation Results

Parameters and True Values	True Utility	Monetary Payoff	Incorrect Utility	Parametric Utility	Non-parametric Utility
$\rho = 0.8$	0.7997	0.9372	0.8271	0.7975	0.7960
	(0.0088)	(0.0060)	(0.0078)	(0.0119)	(0.0148)
$\phi_0 = \phi_1 = 0.9$	0.8997	0.8906	0.8987	0.8993	0.8993
	(0.0047)	(0.0101)	(0.0046)	(0.0047)	(0.0047)
$\delta_0 = \delta_1 = 0.5$	0.4999	0.3928	0.7206	0.5009	0.5032
	(0.0160)	(0.0131)	(0.0089)	(0.0264)	(0.0344)
$u(1) = 1$	*n.a.*	*n.a.*	*n.a.*	1.0050	1.0159
	n.a.	*n.a.*	*n.a.*	(0.1078)	(0.1519)
$u(4) = 2$	*n.a.*	*n.a.*	*n.a.*	1.9963	1.9990
	n.a.	*n.a.*	*n.a.*	(0.1205)	(0.1558)
$u(9) = 3$	*n.a.*	*n.a.*	*n.a.*	2.9848	2.9820
	n.a.	*n.a.*	*n.a.*	(0.1529)	(0.1916)

Notes: Numbers without parentheses are averages of estimates among 100 Monte Carlo samples. Numbers in parentheses are corresponding standard deviations.

Table 4 shows the simulation results. Again, the misspecification of utility generates considerable bias, especially on δ. Compared with Table 3, with additional restrictions, all parameters are estimated with increased precision, especially for utilities whose standard deviations decrease by more than half.

5 Estimation of Experiments by Chmura et al. (2012)

Selten and Chmura (2008) and Chmura et al. (2012) study a large-scale experiment of matching pennies games. Specifically, they consider 12 games with different payoff matrices, as shown in Table 5. The number in parentheses represents the point a subject receives in the corresponding outcome. The point is exchanged for a monetary reward at a fixed rate (i.e., 1.6 euros per 100 points). Therefore, this paper refers to Table 5 as monetary payoff matrices. Selten and Chmura (2008) and Chmura et al. (2012) classify these 12 games into two types. One is a constant-sum game, as shown in the left column. The other is a non-constant-sum game, as shown in the right column. Importantly, note that this paper distinguishes between monetary reward and utility. Therefore, the left column represents the game that is constant sum in terms of monetary payoff, not necessarily constant sum in terms of actual utility.

The experiment consists of 864 subjects, divided into 54 sessions with 16 subjects per session. Each subject participates in only one session, and each session consists of only 1 of 12 games. Specifically, each constant sum-game was played for 12 sessions, and each non-constant-sum game was run for 6 sessions. Subjects were students at the University of Bonn. In each session, a subject was randomly assigned a player role and played such a role for 200 periods. All subjects knew the monetary payoff matrices of each player, and it was common knowledge. In each period, a subject was randomly matched with another subject in the opposing player role. After each period, the subject was informed about the other player's choice, monetary payoff received, period number and cumulative payoff. An experiment session took around 1.5 to 2 hours with an average earning of about 24 euros.

Selten and Chmura (2008) use this experimental dataset to compare five stationary solution concepts of 2×2 games. Chmura et al. (2012) apply the same experiment to study different learning models, such as reinforcement, self-tuning EWA and impulse matching. As their experiment consists of a large number of subjects, it is an ideal dataset for implementing the identification results in this paper.

For ease of estimation, I consider the parametrized GAL model in Subsection 2.4; this parsimonious version also nests EWA and payoff assessment learning as special cases. In experiments, the behaviors of a fraction of subjects are unresponsive to history. For instance, some individuals simply choose a fixed

Table 5: Monetary Payoff Matrices: Experiments in Selten and Chmura (2008) and Chmura et al. (2012)

Constant-Sum Games				***Non-constant-Sum Games***			
		X	*Y*			*X*	*Y*
Game 1	*X*	[10 , 8]	[0 , 18]	Game 7	*X*	[10 , 12]	[4 , 22]
	Y	[9 , 9]	[10 , 8]		*Y*	[9 , 9]	[14 , 8]
		X	*Y*			*X*	*Y*
Game 2	*X*	[9 , 4]	[0 , 13]	Game 8	*X*	[9 , 7]	[3 , 16]
	Y	[6 , 7]	[8 , 5]		*Y*	[6 , 7]	[11 , 5]
		X	*Y*			*X*	*Y*
Game 3	*X*	[8 , 6]	[0 , 14]	Game 9	*X*	[8 , 9]	[3 , 17]
	Y	[7 , 7]	[10 , 4]		*Y*	[7 , 7]	[13 , 4]
		X	*Y*			*X*	*Y*
Game 4	*X*	[7 , 4]	[0 , 11]	Game 10	*X*	[7 , 6]	[2 , 13]
	Y	[5 , 6]	[9 , 2]		*Y*	[5 , 6]	[11 , 2]
		X	*Y*			*X*	*Y*
Game 5	*X*	[7 , 2]	[0 , 9]	Game 11	*X*	[7 , 4]	[2 , 11]
	Y	[4 , 5]	[8 , 1]		*Y*	[4 , 5]	[10 , 1]
		X	*Y*			*X*	*Y*
Game 6	*X*	[7 , 1]	[1 , 7]	Game 12	*X*	[7 , 3]	[3 , 9]
	Y	[3 , 5]	[8 , 0]		*Y*	[3 , 5]	[10 , 0]

action for the entire 200 periods. I classify these subjects as non-learners, and their behaviors cannot be well explained by a learning model. In this paper, I classify learners as the subjects who pay attention and respond to the other player's past behavior. Specifically, I run an individual-level regression of a subject's current action on his/her past choice, the opponent's past choice, and their interaction. I then select the subject who responds to the opponent's past behavior (i.e., the coefficients of the opponent's past choice and interaction term are jointly significant at the 1% level). The selected subjects are classified as learners and are the focus of estimation. The selection reduces the sample size to 463.

The estimation strategy follows the MLE described in Section 4. It assumes that subjects who participate in different games share the same learning behavior. Therefore, all subjects are pooled together in the estimation.[11] Moreover, as the experiment consists of 19 possible monetary payoff outcomes (i.e.,

[11] As discussed in Subsection 3.4, a mixture model that captures individual heterogeneity is identified. However, specifying utility as an unknown function of money introduces non-concavity even with the restriction of homogeneous agents. A mixture model adds additional non-concavity such that finding the global maximum is extremely burdensome. To ease the estimation burden, I choose to estimate under the homogeneity assumption.

number of distinct points that can be received in 12 games), a fully non-parametric specification of utility function introduces 19 parameters (i.e., utility of each possible monetary reward). This substantially complicates the estimation. To ease the computational burden, I consider a polynomial function of order 4 as an approximation. Additional order increases the log-likelihood by only a relatively small magnitude.

$$u(m) = \beta_1 + \beta_2 m + \beta_3 m^2 + \beta_4 m^3. \tag{10}$$

Table 6 presents the estimated learning parameters for various model specifications. The first column assumes the full model but imposes the restriction that monetary payoff is utility. In comparison, the last column estimates the utility function jointly with learning parameters. The three middle columns present different restrictions imposed in the model. The second column assumes equal weight on the attraction of unchosen and chosen strategies (i.e., $\phi_0 = \phi_1$). The third column estimates the standard EWA model, and the fourth one shows the payoff assessment learning. To make a fair comparison with the full model, the middle three columns also specify utility as an unknown function to be estimated. The comparison excludes the original impulse-matching model in Chmura et al. (2012) because they assume a power probability that is different from the logit probability in this paper. Therefore, the comparison is confronted by a different choice rule. Moreover, if we assume a logit formula for the impulse-matching model, it can be seen as a special case of the EWA model with unknown utility function. As a result, I exclude impulse-matching learning in the comparison. Finally, as all specifications are nested in the full model, a likelihood ratio test can be constructed as a specification test, as shown in the last row. All specifications are rejected against the full model. Furthermore, we reject the null hypothesis that utility is money.

First, ρ is robustly estimated to be zero across all specifications. Note that the estimation imposes no restriction on ρ, as compared with the existing literature, which usually imposes $0 \le \rho \le 1$. The estimate suggests that $N(t) = 1$ for each t. It further implies that ϕ can be interpreted as the fraction of current attraction that is carried to the next period. Such a fraction is constant across time given $\rho = 0$. A comparison between the first and last column confirms the Monte Carlo experiment in Section 4. Specifically, there is a relatively small bias on ϕ but a sizable bias on δ if researchers incorrectly assume monetary payoff as utility. Both δ_0 and δ_1 are downward biased by more than 30%. In addition, the result

Table 6: Estimation Results: Learning Parameters

	Monetary Payoff	Same ϕ $\phi_0 = \phi_1$	EWA $\phi_0 = \phi_1$, $\delta_0 = \delta_1$	Payoff Assessment All 0 But ϕ_1	Full Model
ρ	0.0000	0.0000	0.0000	0	0.0000
	(0.1722)	(0.3201)	(0.2912)	n.a.	(0.0994)
ϕ_0	0.2862***	n.a.	n.a.	0	0.2450***
	(0.0111)	n.a.	n.a.	n.a.	(0.0098)
ϕ_1	0.9058***	0.5976***	0.5889***	0.9606***	0.9164***
	(0.0032)	(0.0072)	(0.0071)	(0.0010)	(0.0031)
δ_0	1.7770***	0.7281***	n.a.	0	2.6031***
	(0.0614)	(0.0206)	n.a.	n.a.	(0.0801)
δ_1	2.2009***	0.7693***	0.7757***	0	3.1679***
	(0.0540)	(0.0166)	(0.0181)	n.a.	(0.0937)
Log-likelihood	-47012	-47117	-47141	-52405	-46141
Specification Test (p-value)	0.0000	0.0000	0.0000	0.0000	n.a.

suggests that $\delta_1 > \delta_0$. It indicates that subjects pay more attention to the action that has a higher forgone utility than the actual utility.

A result that seems striking at first glance is that $\delta_0,\ \delta_1 > 1$. This contrasts with the common wisdom in EWA learning such that $0 \le \delta \le 1$. Based on this result, it is tempting to conclude that subjects pay more attention to the forgone utility than the actual one. However, there exists a more plausible interpretation. In the evolution rule of attraction, an action's weights on its current attraction and the current utility both substantially depend on whether such an action is chosen or not. As the full model shows, subjects put a weight of more than 90% on their chosen strategies' attractions. In contrast, only about 25% of the current attraction is carried into the next period for the unchosen strategy. Instead, subjects' evaluation of such an action is based mainly on its current utility. Equivalently, subjects use mainly past performance to evaluate a chosen action; in contrast, they focus more on current information (i.e., current utility) in the evaluation of the unchosen action. This interpretation is further supported by restricting $\phi_0 = \phi_1$, as in the second column, or the EWA model, in the third column. When $\phi_0 = \phi_1$, players are assumed to forget the attractions of the chosen and unchosen strategies at the same rate. δ is estimated with a downward bias and smaller than 1, which is in line with the literature.

Table 7 presents the estimated utility function under the full model. It is plotted in Figure 2. The blue line represents the estimated utility function, and the dotted line represents the 95% confidence interval. The black line plots the utility when it is assumed to be equal to the monetary payoff.[12] Such an assumption is clearly rejected. The utility function is strictly increasing and concave. It suggests that subjects are risk averse. Moreover, the location $u(0)$ is significantly smaller than zero. It indicates that subjects view an action with zero monetary reward as unsuccessful. This is plausible given that each subject expects to earn a positive reward in order to participate in the experiment. It also provides empirical evidence such that assuming $u(0) = 0$ is a restriction, rather than a normalization.

Table 7: Estimation Results: Utility Function

Constant	-0.0738^{***} (0.0122)
m	0.1108^{***} (0.0115)
m^2	-0.0075^{***} (0.0008)
m^3	0.0002^{***} (0.00002)

A common concern about a more general model is that it may be over fitted. To evaluate this potential problem, I compare the out-of-sample predictions of various models. First, I estimate the model based on the constant-sum game only and use the estimated results to predict subjects' behaviors in the non-constant-sum game. Second, I reverse the above procedure and predict the constant-sum game. I choose out-of-sample log-likelihood as the measure of predictability. For both constant-sum and non-constant-sum games, the full model performs substantially better than any restricted model in out-of-sample predictions, as shown in Figure 3. Furthermore, the increase of out-of-sample log-likelihood is at a similar level as the increase of in-sample log-likelihood. Moreover, other nested models, regardless of restrictions on learning parameters or utility function, have similar out-of-sample predictability.[13]

[12] The black line is scaled down by the sensitivity parameter. Specifically, the plot is $u(m) = \lambda m$ where λ is estimated.

[13] Payoff assessment learning is excluded from this comparison as it performs substantially worse than any model.

Figure 2: Utility Function

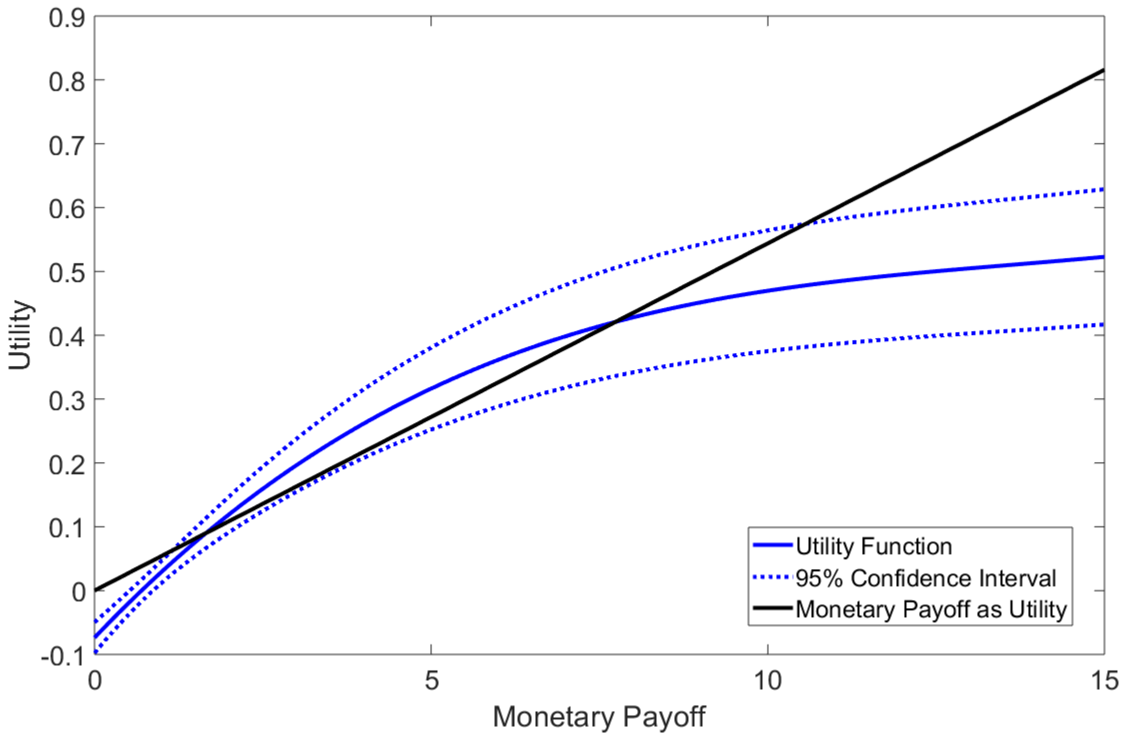
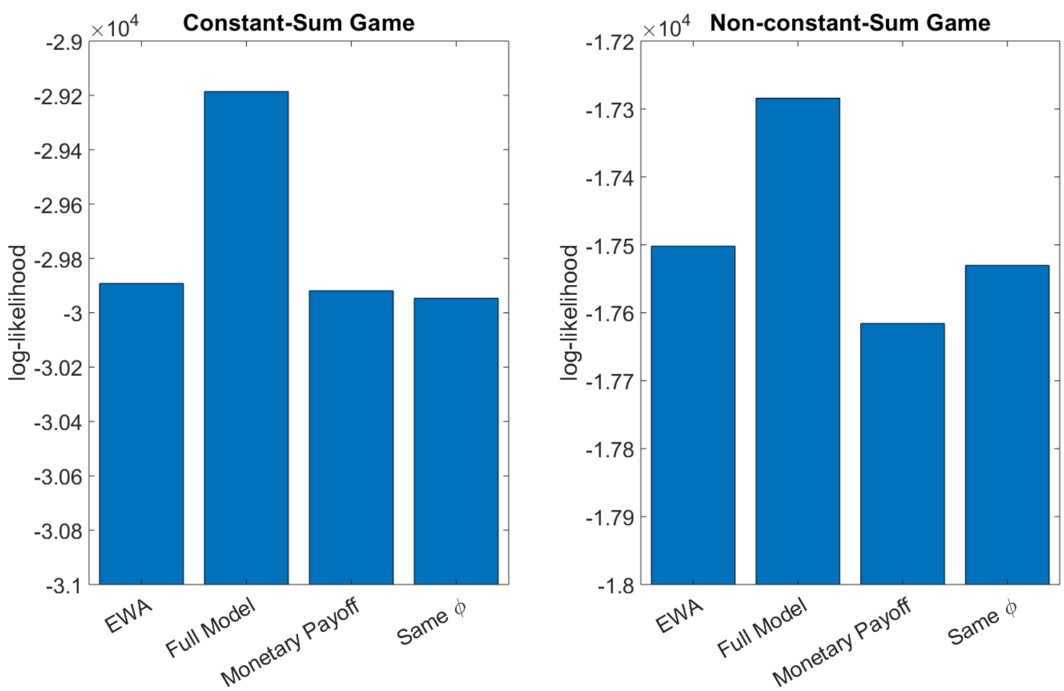


Figure 3: Out-of-sample Log-likelihood



6 Conclusion

This paper focuses on the identification and estimation of adaptive learning models. I first propose a generalized learning rule that allows the weights on both an action's current attraction and current utility to depend on whether such an action is chosen or not. This nests commonly used learning rules as special cases. In addition, I show that adding a constant to the utility function induces different choice dynamics for a wide range of parameter values; such a range includes a broad class of learning models, such as EWA, payoff assessment and impulse matching. This property provides an additional reason, besides the concern regarding risk preference, for the misspecification of utility as monetary payoff. Second, this paper proposes estimating the utility function jointly with the learning parameters. I show that both the utility and the generalized adaptive learning model are point identified under weak conditions. Finally, using an experimental dataset by Selten and Chmura (2008) and Chmura et al. (2012), I reject the null hypothesis that monetary payoff is utility. Incorrectly imposing such an assumption generates considerable bias for both in-sample estimation and out-of-sample prediction.

References

Bracht, J. and Ichimura, H. (2001). Identification of a general learning model on experimental game data. *Working Paper*.

Brown, G. W. (1951). Iterative solution of games by fictitious play. *Activity Analysis of Production and Allocation*. John Wiley & Sons, New York.

Cabrales, A. and Garcia-Fontes, W. (2000). Estimating learning models from experimental data. *Working Paper*.

Camerer, C. and Ho, T. H. (1999). Experience-weighted attraction learning in normal form games. *Econometrica*, 67(4):827–874.

Camerer, C. F., Ho, T. H., and Chong, J. K. (2002). Sophisticated experience-weighted attraction learning and strategic teaching in repeated games. *Journal of Economic Theory*, 104(1):137–188.

Cheung, Y. W. and Friedman, D. (1997). Individual learning in normal form games: Some laboratory results. *Games and Economic Behavior*, 19(1):46–76.

Chmura, T., Goerg, S. J., and Selten, R. (2012). Learning in experimental 2×2 games. *Games and Economic Behavior*, 76(1):44–73.

Chmura, T., Goerg, S. J., and Selten, R. (2014). Generalized impulse balance: An experimental test for a class of 3× 3games. *Review of Behavioral Economics*, 1(1):27–53.

Cominetti, R., Melo, E., and Sorin, S. (2010). A payoff-based learning procedure and its application to traffic games. *Games and Economic Behavior*, 70(1):71–83.

Erev, I. and Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review*, 88(4):848–881.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition and cooperation. *Quarterly Journal of Economics*, 114(5):817–868.

Feltovich, N. (2000). Reinforcement-based vs. belief-based learning models in experimental asymmetric-information games. *Econometrica*, 68(3):605–641.

Feri, F., Irlenbusch, B., and Sutter, M. (2010). Efficiency gains from team-based coordination: Large-scale experimental evidence. *American Economic Review*, 100(4):1892–1912.

Fudenberg, D. and Levine, D. K. (1995). Consistency and cautious fictitious play. *Journal of Economic Dynamics and Control*, 19(5-7):1065–1089.

Fudenberg, D. and Levine, D. K. (1998). *The Theory of Learning in Games*. MIT Press, Cambridge.

Goeree, J. K., Holt, C. A., and Palfrey, T. R. (2003). Risk averse behavior in generalized matching pennies games. *Games and Economic Behavior*, 45(1):97–113.

Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, 3(4):367–388.

Harrison, G. W. and Rütstrom, E. E. (2008). Risk aversion in the laboratory. *Research in Experimental Economics*, 12(12):41–196.

Heinemann, F., Nagel, R., and Ockenfels, P. (2009). Measuring strategic uncertainty in coordination games. *The Review of Economic Studies*, 76(1):181–221.

Ho, T. H., Camerer, C. F., and Chong, J. K. (2007). Self-tuning experience weighted attraction learning in games. *Journal of Economic Theory*, 133(1):177–198.

Kahneman, D., Knetsch, J., and Thaler, R. (1986). Fairness and the assumptions of economics. *Journal of Business*, 59(4):285–300.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–293.

McKelvey, R. D. and Palfrey, T. R. (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior*, 10(1):6–38.

Mookherjee, D. and Sopher, B. (1994). Learning behavior in an experimental matching pennies game. *Games and Economic Behavior*, 7(1):62–91.

Mookherjee, D. and Sopher, B. (1997). Learning and decision costs in experimental constant sum games. *Games and Economic Behavior*, 19(1):97–132.

Ochs, J. (1995). Games with unique, mixed strategy equilibria: An experimental study. *Games and Economics Behavior*, 10(1):202–217.

Roth, A. E. and Malouf, M. W. (1979). Game-theoretic models and the role of information in bargaining. *Psychological Review*, 86(6):574–594.

Salmon, T. C. (2001). An evaluation of econometric models of adaptive learning. *Econometrica*, 69(6):1597–1628.

Sarin, R. and Vahid, F. (1999). Payoff assessments without probabilities: A simple dynamic model of choice. *Games and Economic Behavior*, 28(2):294–309.

Sarin, R. and Vahid, F. (2001). Predicting how people play games: A simple dynamic model of choice. *Games and Economic Behavior*, 34(1):104–122.

Selten, R. and Chmura, T. (2008). Stationary concepts for experimental 2×2 games. *American Economic Review*, 98(3):938–966.

Van Huyck, J. B., Battalio, R. C., and Beil, R. O. (1990). Tacit coordination games, strategic uncertainty, and coordination failure. *American Economic Review*, 80(1):234–248.

Wilcox, N. T. (2006). Theories of learning in games and heterogeneity bias. *Econometrica*, 74(5):1271–1292.

A Appendix

A.1 Conditions on Monetary Payoff Matrix

Define the following matrix:

$$\mathbf{M}_i = \begin{pmatrix} m_i(s_i^1, \mathbf{s}_{-i}^1), & m_i(s_i^1, \mathbf{s}_{-i}^1), & \cdots & m_i(s_i^2, \mathbf{s}_{-i}^1), & \cdots & m_i(s_i^{J_i-1}, \mathbf{s}_{-i}^{\prod_{i' \neq i} J_{i'}}) \\ m_i(s_i^2, \mathbf{s}_{-i}^1), & m_i(s_i^3, \mathbf{s}_{-i}^1), & \cdots & m_i(s_i^3, \mathbf{s}_{-i}^1), & \cdots & m_i(s_i^{J_i}, \mathbf{s}_{-i}^{\prod_{i' \neq i} J_{i'}}) \end{pmatrix}. \tag{11}$$

A single column in $\mathbf{M}_i$ represents the monetary payoffs of two player i's strategies given an action profile chosen by other players. Therefore, $\mathbf{M}_i$ represents the monetary payoffs of all pairs of two actions for any possible choice of other players. This matrix then measures player i's choice incentive in the game. Denote $\mathbf{M} = (\mathbf{M}_1, \mathbf{M}_2, \cdots, \mathbf{M}_n)$ as the matrix of monetary payoffs for all players. Let $\mathbf{m}_l$ be the l^{th} column of $\mathbf{M}$. Assumption 3 states the conditions on $\mathbf{M}$ that achieve the identification of utility when $\phi(t,0) = \phi(t,1)$.

Assumption 3. *Matrix* $\mathbf{M}$ *contains two pairs of columns, one pair denoted by* $\mathbf{m}_l$, $\mathbf{m}_{l'}$ *and another pair by* $\mathbf{m}_k$, $\mathbf{m}_{k'}$*, such that:*[14]

(a) $\mathbf{m}_l$ *and* $\mathbf{m}_{l'}$ *share one and only one identical element denoted by* m^1*; similarly,* $\mathbf{m}_k$ *and* $\mathbf{m}_{k'}$ *share one and only one identical element denoted by* m^2.[15]

(b) If $m^1 = \max(\mathbf{m}'_l, \mathbf{m}'_{l'})$*, then* $m^2 \neq \max(\mathbf{m}'_k, \mathbf{m}'_{k'})$.

(c) If $m^1 = \min(\mathbf{m}'_l, \mathbf{m}'_{l'})$*, then* $m^2 \neq \min(\mathbf{m}'_k, \mathbf{m}'_{k'})$.

Although Assumption 3 may be difficult to interpret at first glance, it imposes weak restrictions on monetary payoffs. Consider a game in which some players have more than two choices. Moreover, player i has three actions that generate distinct monetary payoffs for some of the other players' choices; specifically, denote these three monetary rewards by (m', m'', m'''). Assumption 3 is satisfied by considering two pairs, (m', m'') with (m', m''') and (m', m'') with (m'', m'''). Consequently, Assumption 3 holds for a broad class of games in which some players have more than two actions. An important class of games that is excluded from Assumption 3 is binary-choice symmetric games (i.e., $\mathbf{M}_i$ equals $\mathbf{M}_{i'}$ up to the permutation

[14] $\mathbf{m}_l$ and $\mathbf{m}_k$ can be either the same or different.

[15] m^1 and m^2 can be either the same or different.

for any player i and i'). However, the matrix $\mathbf{M}$ can be extended to multiple treatments in which each treatment shares a different monetary payoff matrix; then Assumption 3 provides guidance on how to design multiple treatments in a symmetric binary-choice game to achieve the identification.

A.2 Proofs

Proof of Proposition 1. Express equation (1) recursively as a weighted sum of initial attraction and s_i^j's utility in each period.

$$A_{i,t+1}^j = \begin{cases} \frac{\phi[1,\mathbb{1}(s_i(1)=s_i^j)]}{N(1)} A_{i,1}^j + \frac{f(s_i^j,\mathbf{s}(1))}{N(1)} \pi_i[s_i^j, \mathbf{s}_{-i}(1)], \text{ if } t = 1, \\ \frac{\prod_{l=1}^t \phi[t,\mathbb{1}(s_i(l)=s_i^j)]}{\prod_{l=1}^t} A_{i,1}^j + \frac{f(s_i^j,\mathbf{s}(t))}{N(t)} \pi_i[s_i^j, \mathbf{s}_{-i}(t)] + \sum_{\tau=1}^{t-1} \frac{\prod_{l=\tau+1}^t \phi[l,\mathbb{1}(s_i(l)=s_i^j)] f(s_i^j,\mathbf{s}(l))}{\prod_{l=\tau}^t N(\tau)} \pi_i[s_i^j, \mathbf{s}_{-i}(\tau)], \text{ if } t \geq 2. \end{cases} \tag{12}$$

Under the logit formula, the choice probability is determined by the difference of attractions for any pair of actions. For any two strategies s_i^j and s_i^k, the difference of attractions $A_{i,t+1}^j - A_{i,t+1}^k$ is

$$\begin{cases} \frac{f(s_i^j,\mathbf{s}(1))}{N(1)} \pi_i[s_i^j,\mathbf{s}_{-i}(1)] - \frac{f(s_i^k,\mathbf{s}(1))}{N(1)} \pi_i[s_i^k,\mathbf{s}_{-i}(1)], \text{ if } t=1, \\ \frac{f(s_i^j,\mathbf{s}(t))}{N(t)} \pi_i[s_i^j,\mathbf{s}_{-i}(t)] - \frac{f(s_i^k,\mathbf{s}(t))}{N(t)} \pi_i[s_i^k,\mathbf{s}_{-i}(t)] + \sum_{\tau=1}^{t-1} \frac{\prod_{l=\tau+1}^t \phi[l,\mathbb{1}(s_i(l)=s_i^j)] f(s_i^j,\mathbf{s}(l))}{\prod_{l=\tau}^t N(\tau)} \pi_i[s_i^j,\mathbf{s}_{-i}(\tau)] - \sum_{\tau=1}^{t-1} \frac{\prod_{l=\tau+1}^t \phi[l,\mathbb{1}(s_i(l)=s_i^k)] f(s_i^k,\mathbf{s}(l))}{\prod_{l=\tau}^t N(\tau)} \pi_i[s_i^k,\mathbf{s}_{-i}(\tau)], \text{ if } t\geq 2. \end{cases}$$

Now, let us add a constant $c \neq 0$ to the utility function. Define $\tilde{\pi}_i(\mathbf{s}) = \pi_i(\mathbf{s}) + c$ for any $\mathbf{s} \in \mathcal{S}$. Assuming players' actual utility is $\tilde{\pi}_i(\cdot)$, we can calculate the difference of any two actions' attractions under the new utility function. Let us denote this new difference as $\tilde{A}_{i,t+1}^j - \tilde{A}_{i,t+1}^k$. For the logit formula, the sufficient and necessary condition for choice probability to be invariant for any $c \neq 0$ is

$$(\tilde{A}_{i,t+1}^j - \tilde{A}_{i,t+1}^k) - (A_{i,t+1}^j - A_{i,t+1}^k) = 0.$$

This is equivalent to the following equations:

$$\frac{f(s_i^j,\mathbf{s}(1)) - f(s_i^k,\mathbf{s}(1))}{N(1)} c = 0,$$

$$\frac{f(s_i^j,\mathbf{s}(t)) - f(s_i^k,\mathbf{s}(t))}{N(t)} c + \sum_{\tau=1}^{t-1} \frac{\prod_{l=\tau+1}^t \phi[l,\mathbb{1}(s_i(l) = s_i^j)] f(s_i^j,\mathbf{s}(l)) - \prod_{l=\tau+1}^t \phi[l,\mathbb{1}(s_i(l) = s_i^k)] f(s_i^k,\mathbf{s}(l))}{\prod_{l=\tau}^t N(\tau)} c = 0.$$

The above equations have to hold for any history of action profile, any constant c and any period t. The only condition for the above equations to hold true is when the coefficients on c all equal zero. It implies that $f(s_i^j, \mathbf{s}) = 1$ for all strategy s_i^j and $\mathbf{s}$, and $\phi(t,1) = \phi(t,0)$ for $t > 1$. This completes the proof. $\square$

Proof of Proposition 2. For any period t, consider an arbitrary history denoted by $h_t = (\mathbf{s}(1), \mathbf{s}(2), \cdots, \mathbf{s}(t))$. Similarly, $h_{t-1} = (\mathbf{s}(1), \cdots, \mathbf{s}(t-1))$. Denote the realized outcome at period t as $\mathbf{s}(t) = (s_i^j, \mathbf{s}_{-i})$ for some s_i^j and $\mathbf{s}_{-i}$. Given the evolution of attraction specified by equation (1), we have the following equation for any pair of s_i^j and s_i^k.

$$N(t)[A_{i,t+1}^j(h_t) - A_{i,t+1}^k(h_t)] = \phi(t,1)A_{i,t}^j(h_{t-1}) - \phi(t,0)A_{i,t}^k(h_{t-1}) + \pi_i(s_i^j, \mathbf{s}_{-i}) - f[s_i^k, (s_i^j, \mathbf{s}_{-i})]\pi_i(s_i^k, \mathbf{s}_{-i}). \tag{13}$$

The notation $A_{i,t}^j(h_{t-1})$ emphasizes that the value of attraction depends on the entire history. Such a notation is suppressed for the sake of brevity in the main text. Consider another history h_t', which is identical to h_t except that player i now chooses action s_i^k at period t. It consequently yields an equation similar to equation (13).

$$N(t)[A_{i,t+1}^j(h_t') - A_{i,t+1}^k(h_t')] = \phi(t,0)A_{i,t}^j(h_{t-1}) - \phi(t,1)A_{i,t}^k(h_{t-1}) + f[s_i^j, (s_i^k, \mathbf{s}_{-i})]\pi_i(s_i^j, \mathbf{s}_{-i}) - \pi_i(s_i^k, \mathbf{s}_{-i}). \tag{14}$$

Adding equations (13) and (14) yields

$$N(t)\Big\{[A_{i,t+1}^j(h_t) - A_{i,t+1}^k(h_t)] + [A_{i,t+1}^j(h_t') - A_{i,t+1}^k(h_t')]\Big\} = [\phi(t,0) + \phi(t,1)][A_{i,t}^j(h_{t-1}) - A_{i,t}^k(h_{t-1})] + Q(s_i^j, s_i^k, \mathbf{s}_{-i}),$$

$$\Rightarrow N(t)\Big\{\log\Big[\frac{P_{i,t+1}^j(h_t)}{P_{i,t+1}^k(h_t)}\Big] + \log\Big[\frac{P_{i,t+1}^j(h_t')}{P_{i,t+1}^k(h_t')}\Big]\Big\} = [\phi(t,0) + \phi(t,1)]\log\Big[\frac{P_{i,t}^j(h_{t-1})}{P_{i,t}^k(h_{t-1})}\Big] + Q_1(s_i^j, s_i^k, \mathbf{s}_{-i}), \tag{15}$$

where $Q_1(s_i^j, s_i^k, \mathbf{s}_{-i}) = \pi_i(s_i^j, \mathbf{s}_{-i}) - f[s_i^k, (s_i^j, \mathbf{s}_{-i})]\pi_i(s_i^k, \mathbf{s}_{-i}) + \pi_i(s_i^j, \mathbf{s}_{-i}) - f[s_i^k, (s_i^j, \mathbf{s}_{-i})]\pi_i(s_i^k, \mathbf{s}_{-i})$.

The second line applies the inversion of the logit formula such that $\log[\frac{P_{i,t}^j}{P_{i,t}^k}] = A_{i,t}^j - A_{i,t}^k$. Similarly, consider $t' > t$ and two histories $h_{t'}$ and $h_{t'}'$, which differ only in the action profile at period t'. Specifically, $\mathbf{s}(t') = (s_i^j, \mathbf{s}_{-i})$ for $h_{t'}$ and $\mathbf{s}(t') = (s_i^k, \mathbf{s}_{-i})$ for $h_{t'}'$, respectively. We can derive an equation similar to equation (15).

$$N(t')\left\{\log\left[\frac{P^j_{i,t'+1}(h_{t'})}{P^k_{i,t'+1}(h_{t'})}\right]+\log\left[\frac{P^j_{i,t'+1}(h'_{t'})}{P^k_{i,t'+1}(h'_{t'})}\right]\right\}=[\phi(t',0)+\phi(t',1)]\log\left[\frac{P^j_{i,t'}(h_{t'-1})}{P^k_{i,t'}(h_{t'-1})}\right]+Q(s^j_i,s^k_i,\mathbf{s}_{-i}). \quad (16)$$

Subtracting equation (15) from (16), we get

$$\begin{aligned}&N(t')\left\{\log\left[\frac{P^j_{i,t'+1}(h_{t'})}{P^k_{i,t'+1}(h_{t'})}\right]+\log\left[\frac{P^j_{i,t'+1}(h'_{t'})}{P^k_{i,t'+1}(h'_{t'})}\right]\right\}-N(t)\left\{\log\left[\frac{P^j_{i,t+1}(h_t)}{P^k_{i,t+1}(h_t)}\right]+\log\left[\frac{P^j_{i,t+1}(h'_t)}{P^k_{i,t+1}(h'_t)}\right]\right\}\\ =&[\phi(t',0)+\phi(t',1)]\log\left[\frac{P^j_{i,t'}(h_{t'-1})}{P^k_{i,t'}(h_{t'-1})}\right]-[\phi(t,0)+\phi(t,1)]\log\left[\frac{P^j_{i,t}(h_{t-1})}{P^k_{i,t}(h_{t-1})}\right]. \quad (17)\end{aligned}$$

We then consider another set of histories such that $h''_t=(h_{t-1},(s^j_i,\mathbf{s}'_{-i}))$, $h'''_t=(h_{t-1},(s^k_i,\mathbf{s}'_{-i}))$, $h''_{t'}=(h_{t'-1},(s^j_i,\mathbf{s}'_{-i}))$, and $h'''_{t'}=(h_{t'-1},(s^k_i,\mathbf{s}'_{-i}))$. We can get an equation similar to equation (17).

$$\begin{aligned}&N(t')\left\{\log\left[\frac{P^j_{i,t'+1}(h''_{t'})}{P^k_{i,t'+1}(h''_{t'})}\right]+\log\left[\frac{P^j_{i,t'+1}(h'''_{t'})}{P^k_{i,t'+1}(h'''_{t'})}\right]\right\}-N(t)\left\{\log\left[\frac{P^j_{i,t+1}(h''_t)}{P^k_{i,t+1}(h''_t)}\right]+\log\left[\frac{P^j_{i,t+1}(h'''_t)}{P^k_{i,t+1}(h'''_t)}\right]\right\}\\ =&[\phi(t',0)+\phi(t',1)]\log\left[\frac{P^j_{i,t'}(h_{t'-1})}{P^k_{i,t'}(h_{t'-1})}\right]-[\phi(t,0)+\phi(t,1)]\log\left[\frac{P^j_{i,t}(h_{t-1})}{P^k_{i,t}(h_{t-1})}\right]. \quad (18)\end{aligned}$$

Subtracting equation (17) from (18) yields

$$\begin{aligned}&N(t')\left\{\log\left[\frac{P^j_{i,t'+1}(h''_{t'})}{P^k_{i,t'+1}(h''_{t'})}\right]+\log\left[\frac{P^j_{i,t'+1}(h'''_{t'})}{P^k_{i,t'+1}(h'''_{t'})}\right]-\log\left[\frac{P^j_{i,t'+1}(h_{t'})}{P^k_{i,t'+1}(h_{t'})}\right]-\log\left[\frac{P^j_{i,t'+1}(h'_{t'})}{P^k_{i,t'+1}(h'_{t'})}\right]\right\}\\ =&N(t)\left\{\log\left[\frac{P^j_{i,t+1}(h''_t)}{P^k_{i,t+1}(h''_t)}\right]+\log\left[\frac{P^j_{i,t+1}(h'''_t)}{P^k_{i,t+1}(h'''_t)}\right]-\log\left[\frac{P^j_{i,t+1}(h_t)}{P^k_{i,t+1}(h_t)}\right]-\log\left[\frac{P^j_{i,t+1}(h'_t)}{P^k_{i,t+1}(h'_t)}\right]\right\}. \quad (19)\end{aligned}$$

Since player i's choice probability $P^j_{i,t}$ is observed for any history, equation (19) then identifies $N(t')/N(t)$ for any two t and t'. Moreover, given the normalization that $N(1)=1$, it consequently identifies $N(t)$ for any $t>1$. Term $\phi(t,0)+\phi(t,1)$ can be identified thereafter given the identification of $N(t)$ by equation (17) or (18).

Instead of adding, let us subtract equation (13) from (14). It will yield an equation similar to equation (15), with term $[\phi(t,1)-\phi(t,0)]$ as the coefficient of $\log[\frac{P^j_{i,t}(h_{t-1})}{P^k_{i,t}(h_{t-1})}]$. Applying the same argument as above, we then identify $[\phi(t,1)-\phi(t,0)]$. Therefore, with both $[\phi(t,1)-\phi(t,0)]$ and $[\phi(t,0)+\phi(t,1)]$ identified, we have the identification of $\phi(t,0)$ and $\phi(t,1)$ for any t.

Next, I prove the identification of initial attractions and utility difference. Consider between periods 1 and 2, let $(s_i, \mathbf{s}_{-i})$ be an arbitrary outcome in period 1. For any pair of s_i^j and s_i^k, we can get the following:

$$N(2)\log\left[\frac{P_{i,2}^j(s_i,\mathbf{s}_{-i})}{P_{i,2}^k(s_i,\mathbf{s}_{-i})}\right] = \phi[1,\mathbb{1}(s_i^j = s_i)]A_{i,1}^j - \phi(1,\mathbb{1}(s_i^k = s_i))A_{i,1}^k + f(s_i^j,\mathbf{s})\pi_i(s_i^j,\mathbf{s}_{-i}) - f(s_i^k,\mathbf{s})\pi_i(s_i^k,\mathbf{s}_{-i}). \tag{20}$$

First, consider the case that $\phi(t,0) = \phi(t,1)$. Under this scenario, adding a constant c to the initial attraction of each action has no impact on the difference of any two actions' future attractions. It consequently leaves the sequence of choice probability unchanged. Therefore, a normalization such that $A_{i,1}^1 = 0$ is required for identification. Given such a normalization, initial attraction of any action s_i^j is identified by $A_{i,1}^j = \log[\frac{P_{i,1}^j}{P_{i,1}^1}]$. It consequently yields the identification of $f(s_i^j,\mathbf{s})\pi_i(s_i^j,\mathbf{s}_{-i}) - f(s_i^k,\mathbf{s})\pi_i(s_i^k,\mathbf{s}_{-i})$ for any s_i^j, s_i^k and $\mathbf{s}$ by equation (20).

Suppose now $\phi(t,0) \neq \phi(t,1)$ for some t, the normalization $A_{i,1}^1 = 0$ is no longer innocuous. However, given the argument in the above paragraph, both A_i^j and difference of utilities are identified up to the value of $A_{i,1}^1$. Equivalently, player i's choice probability in any period for any history is a function that depends only linearly on the value of $A_{i,1}^1$. Therefore, player i's choice probability at period $t' > t$ provides information that is sufficient to identify $A_{i,1}^1$. This completes the proof. □

Proof of Proposition 3. Without loss of generality, suppose $\phi(t,0) \neq \phi(t,1)$ when $t = 2$. The case when $\phi(t,0) \neq \phi(t,1)$ for some $t > 2$ is conceptually equivalent and therefore is suppressed. For any action profile $(s_i^j, \mathbf{s}_{-i})$, suppose it is observed in the first two periods; the evolution of attraction $A_{i,t}^j$ is then

$$A_{i,2}^j = \phi(1,1)A_{i,1}^j + u[m_i(s_i^j,\mathbf{s}_{-i})],$$
$$A_{i,3}^j = \frac{\phi(2,1)\phi(1,1)}{N(2)}A_{i,1}^j + \frac{\phi(2,1)}{N(2)}u[m_i(s_i^j,\mathbf{s}_{-i})] + \frac{1}{N(2)}u[m_i(s_i^j,\mathbf{s}_{-i})]. \tag{21}$$

Consider another action of player i, s_i^k. Suppose that $m_i(s_i^k,\mathbf{s}_{-i}) < m_i(s_i^j,\mathbf{s}_{-i})$; note the case when

$m_i(s_i^k, \mathbf{s}_{-i}) \geq m_i(s_i^j, \mathbf{s}_{-i})$ is conceptually equivalent and is suppressed. The evolution of $A_{i,t}^k$ is

$$\begin{aligned} A_{i,2}^k &= \phi(1,0)A_{i,1}^k + \delta_0 u[m_i(s_i^k, \mathbf{s}_{-i})], \\ A_{i,3}^k &= \frac{\phi(2,0)\phi(1,0)}{N(2)}A_{i,1}^k + \frac{\phi(2,0)}{N(2)}\delta_0 u[m_i(s_i^k, \mathbf{s}_{-i})] + \frac{1}{N(2)}\delta_0 u[m_i(s_i^k, \mathbf{s}_{-i})]. \end{aligned} \tag{22}$$

Taking the differences of equations (21) and (22) and applying logit formula, we get

$$\begin{aligned} u[m_i(s_i^j, \mathbf{s}_{-i})] - \delta_0 u[m_i(s_i^k, \mathbf{s}_{-i})] &= \log[\frac{P_{i,2}^j}{P_{i,2}^k}] - \phi(1,1)A_{i,1}^j + \phi(1,0)A_{i,1}^k, \\ (\phi(2,1)+1)u[m_i(s_i^j, \mathbf{s}_{-i})] - (\phi(2,0)+1)\delta_0 u[m_i(s_i^k, \mathbf{s}_{-i})] &= \log[\frac{P_{i,3}^j}{P_{i,3}^k}] - \phi(2,1)\phi(1,1)A_{i,1}^j + \phi(2,0)\phi(1,0)A_{i,1}^k. \end{aligned} \tag{23}$$

Above is a linear equation system with two equations and two unknowns (i.e., $u[m_i(s_i^j, \mathbf{s}_{-i})]$ and $\delta_0 u[m_i(s_i^k, \mathbf{s}_{-i})]$). All other terms are identified given Proposition 2. Therefore, when $\phi(2,0) \neq \phi(2,1)$, the equation system has a unique solution. $u[m_i(s_i^j, \mathbf{s}_{-i})]$ and $\delta_0 u[m_i(s_i^k, \mathbf{s}_{-i})]$ are then identified.

Repeat the above procedure and suppose that $(s_i^k, \mathbf{s}_{-i})$, instead of $(s_i^j, \mathbf{s}_{-i})$, is observed in the first two periods. It then yields the identification of $\delta_1 u[m_i(s_i^j, \mathbf{s}_{-i})]$ and $u[m_i(s_i^k, \mathbf{s}_{-i})]$. The weights δ_1 and δ_0 are identified thereafter given the identification of $u[m_i(s_i^j, \mathbf{s}_{-i})]$ and $u[m_i(s_i^k, \mathbf{s}_{-i})]$. This completes the proof. □

***Proof of Proposition 4*.** Given Assumption 3, consider the following two pairs of vectors:

$$\mathbf{m}_l = \begin{pmatrix} m^1 \\ m^3 \end{pmatrix}, \mathbf{m}_{l'} = \begin{pmatrix} m^2 \\ m^4 \end{pmatrix}, \mathbf{m}_k = \begin{pmatrix} m^2 \\ m^5 \end{pmatrix}, \mathbf{m}_{k'} = \begin{pmatrix} m^2 \\ m^6 \end{pmatrix}.$$

As shown in Proposition 2, the weighted difference of any two actions' utilities are identified. When such a result is applied to vector $\mathbf{m}_l$, the following two terms are identified:

$$\begin{aligned} &u(m^1) - \delta_1 \cdot \mathbb{1}(m^1 \leq m^3) \cdot u(m^3) - \delta_0 \cdot \mathbb{1}(m^1 > m^3) \cdot u(m^3), \text{ when } m^1 \text{ is realized,} \\ &\delta_1 \cdot \mathbb{1}(m^1 \geq m^3) \cdot u(m^1) + \delta_0 \cdot \mathbb{1}(m^1 < m^3) \cdot u(m^1) - u(m^3), \text{ when } m^3 \text{ is realized.} \end{aligned}$$

Similarly, when the same result is applied to all four vectors $\mathbf{m}_l$, $\mathbf{m}_{l'}$, $\mathbf{m}_k$ and $\mathbf{m}_{k'}$, the following eight terms are identified:

$$
\begin{aligned}
&u(m^1) - \delta_1 \cdot \mathbb{1}(m^1 \leq m^3) \cdot u(m^3) - \delta_0 \cdot \mathbb{1}(m^1 > m^3) \cdot u(m^3), \\
&\delta_1 \cdot \mathbb{1}(m^1 \geq m^3) \cdot u(m^1) + \delta_0 \cdot \mathbb{1}(m^1 < m^3) \cdot u(m^1) - u(m^3), \\
&u(m^1) - \delta_1 \cdot \mathbb{1}(m^1 \leq m^4) \cdot u(m^4) - \delta_0 \cdot \mathbb{1}(m^1 > m^4) \cdot u(m^4), \\
&\delta_1 \cdot \mathbb{1}(m^1 \geq m^4) \cdot u(m^1) + \delta_0 \cdot \mathbb{1}(m^1 < m^4) \cdot u(m^1) - u(m^4), \\
&u(m^2) - \delta_1 \cdot \mathbb{1}(m^2 \leq m^5) \cdot u(m^5) - \delta_0 \cdot \mathbb{1}(m^2 > m^5) \cdot u(m^5), \\
&\delta_1 \cdot \mathbb{1}(m^2 \geq m^5) \cdot u(m^2) + \delta_0 \cdot \mathbb{1}(m^2 < m^5) \cdot u(m^2) - u(m^5), \\
&u(m^2) - \delta_1 \cdot \mathbb{1}(m^2 \leq m^6) \cdot u(m^6) - \delta_0 \cdot \mathbb{1}(m^2 > m^6) \cdot u(m^6), \\
&\delta_1 \cdot \mathbb{1}(m^2 \geq m^6) \cdot u(m^2) + \delta_0 \cdot \mathbb{1}(m^2 < m^6) \cdot u(m^2) - u(m^6).
\end{aligned}
$$

Given conditions (b) and (c) in Assumption 3, the above identified terms contain eight unknowns, for instance, δ_1, δ_0 and $u(m^1)$ up to $u(m^6)$. The rank condition is satisfied so all unknowns are identified.

Given the identification of δ_1 and δ_0, consider two arbitrary strategy profiles, $(s_i^j, \mathbf{s}_{-i})$ and $(s_i^k, \mathbf{s}_{-i})$, in which only player i's actions are different. Without loss of generality, assume $m_i(s_i^j, \mathbf{s}_{-i}) > m_i(s_i^k, \mathbf{s}_{-i})$. The following terms are identified:

$$
\begin{aligned}
&u[m_i(s_i^j, \mathbf{s}_{-i})] - \delta_0 u[m_i(s_i^k, \mathbf{s}_{-i})], \\
&\delta_1 u[m_i(s_i^j, \mathbf{s}_{-i})] - u[m_i(s_i^k, \mathbf{s}_{-i})].
\end{aligned}
$$

When at least one of δ_1 and δ_0 is not equal to 1, the rank condition is satisfied. Therefore, both $u[m_i(s_i^j, \mathbf{s}_{-i})]$ and $u[m_i(s_i^k, \mathbf{s}_{-i})]$ are identified. When $\delta_1 = \delta_0 = 1$, the above two terms turn out to be identical, represented as term $u[m_i(s_i^j, \mathbf{s}_{-i})] - \delta_0 u[m_i(s_i^k, \mathbf{s}_{-i})]$, and is identified. This directly implies the identification of the difference of utilities. This completes the proof. $\square$