Another frame, another game? Explaining framing effects in economic games

Philipp Gerlach\textsuperscript{1} \& Bastian Jaeger\textsuperscript{2}

\textsuperscript{1}Center for Adaptive Rationality, Max Planck Institute for Human Development
\textsuperscript{2}Department of Social Psychology, Tilburg University

1. Introduction

In January 1950, Merrill M. Flood and Melvin Dresher put game theory to the test under highly controlled laboratory conditions – probably for the first time in history. The two researchers invited Armen Alchian, a member of the UCLA economics department, and John Williams, the chair of RAND’s math section, and placed them in separate locations of the laboratory. Alchian and Williams independently saw a $2 \times 2$ matrix that listed four payoffs. Each payoff represented a potential combination of the two moves between which they had to choose – simultaneously and incommunicado. The catch was that Alchian’s payoff-maximising option was at the detriment of Williams’ payoffs and vice versa (De Herdt, 2003; Flood, 1958).

Just two months later, Princeton mathematician Albert Tucker outlined to his psychology students the social dilemma faced by Alchian and William. Instead of explaining the dilemma in the plain words used by Flood and Dresher, Tucker invented a tale about two criminal suspects that gave the game its name: the prisoner’s dilemma game (De Herdt, 2003; Field, 2001). But did Tucker – by adding context to the game – indeed explain the same game in other words? Or would Alchian and Williams have made different choices had the instructions they were given taken the form of Tucker’s tale?

It has been repeatedly demonstrated that small changes in the framing of games (i.e., the way in which the game situation is described to participants) can have large effects on players’ choices. For example, referring to a prisoner’s dilemma game as the “Community Game” as opposed to the “Wall Street Game” can double the cooperation rate (Liberman, Samuels, & Ross, 2004). Doubts have therefore arisen as to whether existing theories of human behaviour can account for observations from even the most simplistic and controlled laboratory interactions as represented by games (Hagen & Hammerstein, 2006; Levitt & List, 2007; Weber, Kopelman, & Messick, 2004). This article explains the relevance of framing effects for experimental practice (section 2), reviews the debates about framing effects (section 3), summarises four mechanisms that potentially account for framing effects (section 4), and discusses their scope and grasp (section 5).

2. Relevance

Lacking or conflicting standards for the description of games can threaten experimental replicability. Variation in the descriptions of games may, for example, lead participants to construe the task differently and thus elicit dissimilar choices. Nevertheless, there are surprising differences in experimental practice between and within the two disciplines that make most use of games: experimental economics and psychology.

Experimental economists typically rely on pre-defined scripts to describe games. These scripts are either read aloud or presented to the players in written format. Descriptions frequently focus on the mathematical details of games, and non-mathematical aspects are displayed in plain, abstract and generic language (Binmore, 2010; Camerer & Fehr, 2002, p. 5; Friedman & Sunder, 1994, p. 17). However, there are no explicitly defined standards for the language used for the non-mathematical aspects. For example, players may be addressed as “you and the other”, moves labelled as “A” and “B” and the game referred to as “the situation”. Alternatively, labels may be taken from the domain of games (e.g., “the players”, “the game”), finance (e.g., “the buyer”, “the seller”, “the exchange”) or experimental practice (e.g., “the participants”, “the experiment”). Usually, players are not informed about the purpose of the game or given any real-world examples before making their decisions. Some experimental economics textbooks even directly advise researchers against including...
any real-world references beyond the game itself (Camerer & Fehr, 2002; Friedman & Sunder, 1994, p. 17), suggesting that only a generic and abstract language is “neutral” and permits “clean assessment” (Engel & Rand, 2014, p. 387). Overall, stripping away context is a widely accepted and even recommended research practice among economists (Holt, 2006, p. 13).

Psychologists, in contrast, have less binding conventions for the implementation of games. It is common to ad-lib descriptions and not to provide all players with exactly the same description. Instead descriptions may vary from experiment to experiment — even from session to session (Hertwig & Ortmann, 2001). Similarly, references to real-life events are not proscribed. Psychologists sometimes embed games within coherent stories (e.g., C. J. Orwant & Orwant, 1970) or instruct players to imagine that the game represents a specific situation, such as a conflict between business owners or nation states (e.g., Eiser & Bhavnani, 1974; Handgraaf, Dijk, Vermunt, Wilke, & Dreu, 2008).

In parallel to these different experimental set-up practices, choice theories in economics and psychology differ in terms of whether they account for how tasks are described (Zwick, Erev, & Budescu, 1999). Most choice theories in economics do not account for the framing of games. Indeed, description invariance is a tenet of the widely used expected utility framework: the same information about alternatives is assumed to lead to the same choices regardless of its framing. In this line of thinking, language is the “surface structure”, whereas its logical, mathematical rules provide the “deep structure” according to which players decide (Cooper & Kagel, 2003; Wagenaar, Keren, & Lichtenstein, 1988). A notable exception is valence framing, a term which in economics is frequently used synonymously with framing (we return to the integration of valence framing into economic theories in section 4.1).

In psychology, by contrast, modelling the effect of framing has become an increasingly popular research topic (Druckman, 2001; Elliott & Hayward, 1998; Levin, Schneider, & Gaeth, 1998). In their seminal studies, Kahneman and Tversky for example argued that framing can strongly influence judgements between treatment conditions, even causing preference reversals (Kahneman & Tversky, 1979, 1984; Tversky & Kahneman, 1981, 1986). In psychology, framing is rather broadly defined and refers to the provision of differently worded but logically equivalent situations to an audience; we use the term framing in this way throughout this article.

3. Frames and games

There have been three major scientific debates about framing effects in experimental games. The first focuses on a specific class of games: coordination games (e.g., matching pennies, left/right game). In coordination games, multiple players choose simultaneously and — unlike in the prisoner’s dilemma — making (dis-)similar choices results in the highest payoffs for all players. There are no conflicting interests involved. Different frames can refer players to a unique but principally arbitrary move. In other words, frames can induce different focal points because players use the information provided to coordinate their choices (Bacharach & Bernasconi, 1997; Colman, Pulford, & Rose, 2008; Schelling, 1960).

The second debate focuses on the valence displayed. Valence framing (in economics often simply: framing) refers to a class of situations in which the same information is presented as either losses or gains (Levin et al., 1998). In economic games, valence framing involves changes in the suggested property rights. For instance, common pool games can be described as decisions on how much to “give” to a public good (public goods games) or how much to “take” from a collectively shared resource (commons dilemmas). Although both games have the same payoff structure, they are psychologically not the same (Andreoni, 1990, 1995). Depending on the valence frame, choices vary systematically (Andreoni, 1995; e.g., Brewer & Kramer, 1986; Cookson, 2000; Dijk & Wilke, 2000; Dufwenberg, Gächter, & Hennig-Schmidt, 2011; Fleishman, 1988; McCusker & Carnevale, 1995; McDaniel & Sistrunk, 1991; Park, 2000; Sonnemans, Schram, & Offerman, 1998; Willinger & Ziegelmeier, 1999). Researchers have already identified a host of factors which determine whether the give-frame or the take-frame leads to higher contributions in common pool games, including its interaction with the payoffs (Brandts & Schwieren, 2007) and the participant pool (Goerg & Walkowitz, 2010). Overall, valence framing effects in games are well studied (for meta-studies that consider valence framing effects, see Engel, 2011, for dictator games; Sally, 1995, for prisoner’s dilemmas; Zelmer, 2003, for public goods games).
The third debate revolves around the provision of differently worded but logically equivalent descriptions of otherwise unchanged games (i.e., the psychological definition of framing). This debate is the youngest (Ellingsen, Johannesson, Mollerstrom, & Munkhammar, 2012; Hagen & Hammerstein, 2006; Levitt & List, 2007; Weber et al., 2004) — although psychologists (Deutsch, 1957, 1958, 1960) and experimental economists (Hoffman & Spitzer, 1985; Hoffman, McCabe, & Smith, 1996) have long known that framing effects are not limited to either changes in valence or to coordination games. For instance, the instructions to non-cooperative games, in which players choose independently and no third party can enforce cooperation, can emphasise either the competitive nature of the situation or the possible group advantage of cooperation. Indeed, an emphasis on one aspect or the other has been identified to change the cooperation rate in prisoner’s dilemmas as much as from 13% to 89% (Deutsch, 1957, 1958, 1960). In a meta-analysis, the emphases were identified as two of the three most powerful predictors of choices (Sally, 1995). To distinguish the psychological definition of framing from valence framing, we suggest calling these changes in the wording of games context framing. Context framing shapes players’ experience of the game by (1) associating the game with different entities and/or (2) stressing specific aspects of the game. Examples of context framings include changing the title of the game (e.g., “Community Game” vs. “Wall Street Game”; Liberman et al., 2004), embedding the game in a coherent story (e.g., by referring to local sharing rituals that resemble the structure of the game; Lesorogol, 2007), and/or stressing the dependency of the other player in non-interactive games (e.g., by adding the sentence “Note that your recipient relies on you” to the instructions; Brañas-Garza, 2007).

4. Explaining framing effects

4.1. Valence framing effects

Valence framing effects are often explained by prospect theory, which argues that gains and losses are evaluated according to different, nonlinear value functions (Kahneman, Knetsch, & Thaler, 1990; Knetsch, 1989; Kühberger, 1998). The loss function is generally steeper than the gain function, describing that the same amount is weighted more heavily when lost than when gained. Notably, prospect theory is increasingly accepted by economists, and efforts have been made to integrate it into existing theories on game play (Dijk & Wilke, 2000; Iturbe-Ormaetxe, Ponti, Tomás, & Ubeda, 2011). A prominent example is the work by Andreoni (1995), who suggested that players associate positive externalities of their own actions with a ‘warm glow’ and negative externalities with a ‘cold prickle’. In public goods games, for example, the positive give-frame stresses positive externalities arising from one’s contribution; in commons dilemmas, in contrast, the take-frame stresses negative externalities arising from one’s withdrawal. Because these externalities are weighted according to different value functions, the give-frame and the take-frame may lead to different choices. Yet despite Andreoni’s attempt to explain valence framing effects in games, the debate remains unsettled (Gächter, Kölle, & Quercia, 2014; Keysar, Converse, Wang, & Epley, 2008; Schwartz-Shea & Simmons, 1995). In common pool games, for example, valence framing effects depend on several boundary conditions (see section 3), and they trigger not only different value functions and choices but entire strategies (Gächter et al., 2014).

4.2. Context framing effects

There is even more substantial disagreement on how context framing effects should be explained. In the following, we briefly summarise some of the most popular theories and categorise them to one of three classes. Each class argues for a different psychological mechanism underlying context framing effects. Notably, the proposed explanations are not mutually exclusive but may hold at the same time – and they may also explain valence framing effects.
4.2.1. Coordination device theories

The most common explanation for cooperation and non-zero transfers is provided by social preference theories. Social preference theories assume that at least some players take into account not only their own payoffs, but also the payoffs of others (e.g., Andreoni, Brown, & Vesterlund, 2002; Levine, 1998; Loewenstein, Thompson, & Bazerman, 1989) or the relation between their own and others’ payoffs (e.g., Bolton, 1991; Bolton & Ockenfels, 2000; Delue, McClintock, & Liebrand, 1993; Fehr & Schmidt, 2006; Liebrand & McClintock, 1988). In the Allportian tradition, social preferences are frequently assumed to be stable characteristics of the player that are faithfully expressed across time and contexts (Camerer & Thaler, 1995; Loomes, 1999). Therefore, social preference theories cannot directly account for context framing effects.

However, social preference theories can indirectly account for context framing effects if first-order beliefs (FOBs) are taken into account (Camerer & Fehr, 2002, footnote 6; Dreber, Ellingsen, Johannesson, & Rand, 2013; Ellingsen et al., 2012). FOBs are the player’s expectations about her partner’s choices. In the absence of more reliable knowledge (e.g., via mutual promises), players may base their FOBs on the context frame. Hence, context frames can induce different choices among players, who choose according to their FOBs. For example, conditionally cooperative players prefer cooperation to defection in prisoner’s dilemmas but only if they believe that their partner will cooperate (see Battigalli & Dufwenberg, 2009; Cubitt, Drouvelis, & Gächter, 2011; Fischbacher, Gächter, & Fehr, 2001). Framing a prisoner’s dilemma as “the Community Game” could thus increase a conditionally cooperative player’s FOB in cooperation if she believes that it also raises the FOB of her partner, assuming that his choice is similarly FOB-dependent. As a consequence, the context framing makes the conditionally cooperative player choose differently without changing her preferences. In this sense, FOBs can become self-fulfilling, and context frames may serve as coordination devices for players with conditional social preferences (similar to focal points: see Schelling, 1960).

Coordination device theories account for context framing effects in games in which players choose simultaneously and without knowledge about the other player, such as prisoner’s dilemmas. However, they do not account for context framing effects in sequential move games (Ellingsen et al., 2012). A prominent example is the trust game that consists of two players and two stages (Berg, Dickhaut, & McCabe, 1995). In stage one, trustees can invest any part of their endowment in the trustee. The investment is then multiplied. In stage two, trustees can decide to return any of the multiplied amount to the trustor. Because trustees choose after they know the trustor’s choice (direct response) or because trustees make conditional choices for each of the trustor’s potential choices (strategy method), trustees need not infer the trustor’s FOBs from the context frame: they either know the trustor’s choice (direct response) or they can modify their responses based on all potential choices of the trustor (strategy method). Because trustors will anticipate that context frames have no impact on the trustee’s FOBs and hence her choice, the trustor chooses independently of the context frame. Yet context framing effects have nevertheless been observed in trust games (e.g., Al-Ubaydli, Houser, Nye, Paganelli, & Pan, 2013; Cronk & Wasielewski, 2008).

4.2.2. Group identity theories

A competing class of theories assumes that social preferences themselves are flexible and to some degree dependent on the context frame (e.g., Akerlof & Kranton, 2000; Bacharach, 1999, 2006; Bacharach & Bernasconi, 1997; Brewer & Kramer, 1986; Gold & Sugden, 2007a, 2007b; Kramer & Brewer, 1984; Sugden, 1993, 2000, 2015; Wit & Wilke, 1992). Foremost, group identity theories assume that players first categorise all players, including themselves, into groups. They then either identify with a group (i.e., other players belong to one’s ingroup) or do not identify with a group (i.e., other players belong to the outgroup). Here, context frames can direct attention and initiate group identification among players. For example, an emphasis on players’ common goals might elicit ingroup identification and prompt players to shift their preferences up to the group level. In this case, players aim at maximising joint payoffs. Alternatively, an emphasis on conflicting goals or opposing worldviews might lead to the perception of other players as members of the outgroup. If players enter the game with a default social preference for maximising joint payoffs, an emphasis on conflicting goals can prompt players to shift their social preferences down and cause them to maximise their own payoffs rather than joint payoffs.
Group identity theories can account for context framing effects in variable-sum games in which the total sum of the payoffs depends on players’ individual choices. In these games, players’ choices are located on a continuum ranging from maximising their individual payoffs only to maximising the group’s joint payoffs. Prisoner’s dilemmas, public goods games and trust games are all examples of variable-sum games. However, group identity theories do not apply to games in which the total payoffs are fixed (e.g., dictator games). Yet context framing effects have nevertheless been observed in fixed-sum games (e.g., DeScioli & Krishna, 2013; Lesorogol, 2007).

### 4.2.3. Social norms theories

The third class of theories argues that context frames induce players to associate games with different social contexts that are, in turn, associated with different social norms. Social norms may elicit different choices because (at least some) players make rule-based decisions (Bardsley, 2010; Bettenhausen & Murnaghan, 1991; Bicchieri, 2006; Bicchieri & Xiao, 2009; Bicchieri & Zhang, 2012; Biel & Thøgersen, 2007; Borgstede, Dahlstrand, & Biel, 1999; Elster, 1989; Kallgren, Reno, & Cialdini, 2000; Young, 2003). In rule-based decision-making, a player first categorises the game as an exemplar of a class of social situations (e.g., this game resembles situations of class A) and then associates behavioural rules with that class of social situations (in situations of class A, do X). Social norms theories assume that people prefer to choose according to the associated rule if that rule is in line with their FOBs and/or second-order beliefs (SOBs). That is, the player believes the other(s) would follow the same rule (FOB); and/or the player believes the other(s) expect(s) her to make a specific choice (SOB). Context frames may change the categorisation of games, thereby evoking different rule associations, which could in turn lead to different FOBs and/or SOBs. Players then decide by matching their choice with their FOB and/or SOB.

Social norms theories can account for context framing effects without restriction to a particular game: In all games, a player’s rule association and hence her FOBs and/or SOBs can depend on the context frame. At the same time, social norms theories do not specify whether the resulting context framing effect is due to FOBs or SOBs. This is problematic for experimental games, because the player’s SOBs are not necessarily limited to the other player but can include SOBs about the experimenter. The latter would mean that players deduce from the experimenter’s framing what the experimenter expects from the player. This effect is known as an experimenter demand effect (Bardsley, 2008; Ortmann, 2005; also Sher & McKenzie, 2006, 2008; Zizzo, 2010). Thus, although social norms theories capture context framing effects in all games, they tend to remain vague about the specific mechanisms: Do players choose according to their FOBs or their SOBs? In the latter case, are the SOBs about the other player or the experimenter?

### 5. Discussion

Framing effects in games are an empirically well-studied phenomenon. However, a coherent theoretical explanation of the observed effects is still lacking. This article distinguished between two types of framings — valence framing and context framing — and provided an overview of three general classes of theories that may account for the observed changes in behaviour. All classes of theories have their limitations: prospect theory applies only to valence framing and cannot capture the changes observed in strategy selection; coordination device theories and group identity theories both apply to specific games only, although framing effects have also been observed in other games; social norms theories remain vague about the underlying mechanisms and leave open whether framing effects are experimenter demand effects. At the moment, theory-driven investigations are the exception (e.g., Ellingsen et al., 2012; Gächter et al., 2014). Clearly, more research is required to clarify the mechanisms underlying framing effects.
6. References


new approaches to solving real-world problems (pp. 175–180). Berlin, Germany: Springer.


Sally, D. (1995). Conversation and cooperation in social dilemmas: A meta-analysis of experiments from...


