Journal of Behavioral and Experimental Finance 23 (2019) 189–197

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Finance

journal homepage: www.elsevier.com/locate/jbef

Full length article

Perception of intentionality in investor attitudes towards financial risks

Peter Bossaerts [a,b,*], Shinsuke Suzuki [c,d], John P. O'Doherty [c,e]

[a] Brain, Mind and Markets Laboratory, Department of Finance, Faculty of Business and Economics, The University of Melbourne, Parkville, VIC 3010, Australia
[b] The Florey Institute of Neuroscience and Mental Health, Parkville, VIC 3010, Australia
[c] Division of The Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125, USA
[d] Frontier Research Institute for Interdisciplinary Sciences, Tohoku University, Sendai, 980-8578, Japan
[e] Computation and Neural Systems, California Institute of Technology, Pasadena, CA 91125, USA

ARTICLE INFO

Article history:
Received 8 December 2017
Accepted 27 December 2017
Available online 11 January 2018

Keywords:
Financial risk perception
Intentional stance
Theory of Mind
Experimental finance
Trading skills

ABSTRACT

Traditionally, financial market participation has been treated as analogous to playing games of chance with a physical device such as roulette. Here, we propose that humans treat financial markets as intentional agents, with own beliefs and aspirations. As a result, the capacity to infer the intentions of others, Theory of Mind, explains behaviour. As evidence, we appeal to results from recent studies of: (i) forecasting in the presence of insiders, (ii) trading in markets with bubbles, and (iii) financial contagion. Intensity of, and skill in, Theory of Mind explains heterogeneity, not only in choices but also in neural activation.

> The Fed has ignored Mr Market for a very long time and he has felt neglected and marginalised.
>
> Do not be too surprised if his anger upends the best laid plans of mice and Fed.
>
> *Financial Times, 21 January 2016*

1. Motivation

Theoretical and empirical analyses of trading and pricing in financial markets have so far assumed that investors treat financial risks as non-intentional, as if generated by a physical device the workings of which satisfy blind laws of nature, like a roulette wheel (Fig. 1(a)). As a result, decision theory and machine learning, disciplines that focus on physical risks, have formed the basis of investments analysis (Markowitz et al., 2000), asset pricing theory (Radner, 1972), and more recently, exploration of the neurobiological foundations of financial decision-making (Bossaerts, 2009). Sub-optimal choices and ensuing mis-pricing have been explained in terms of heuristics that humans are known to resort to when playing games of chance in casinos, such as gamblers' and hot hand fallacies, probability distortion, or disposition effects (Hirshleifer, 2001).

We propose instead that humans tend to personify financial risks, with which we mean that they endow them with *intentionality*. Sub-optimal behaviour then emerges as the result of mistaken application of *Theory of Mind*, i.e., the capacity (of mostly higher primates) to put oneself in the feet of another person, perceive this person's beliefs, desires, and hence, intentions, and to act on these perceptions (Frith and Frith, 2005). Theory of Mind is best exemplified in instances of the matching pennies game, such as when a soccer player tries to score in penalty kicks by avoiding shooting the ball in the corner that the goal keeper chooses to jump to (the penalty kicker is rewarded for mismatching, while the goal keeper wins when she matches; see Fig. 1(B).)

Financial risks are generated in financial markets; those markets cannot and should not be thought of as intentional. At best, intentionality is indirect, when they are populated with intentional agents. Order flow and pricing are consequences of the meeting of those agents, and as is well known from social choice theory, actions that emerge from group decision-making generally cannot be represented as if made by a single (representative) agent, exceptions notwithstanding (Sen and Pattanaik, 1969). Yet humans appear to have a remarkable capacity to attribute beliefs and desires to objects and systems that cannot possibly have those. This capacity often helps them understand and better predict outcomes. That is, humans often take an *intentional stance* (Dennett, 1989). Here, we will show that this is true for financial risks as well.

* Corresponding address: Brain, Mind and Markets Laboratory, Department of Finance, Faculty of Business and Economics, The University of Melbourne, 198 Berkeley Street, 3010 Carlton (VIC), Australia.
E-mail address: peter.bossaerts@unimelb.edu.au (P. Bossaerts).

https://doi.org/10.1016/j.jbef.2017.12.011

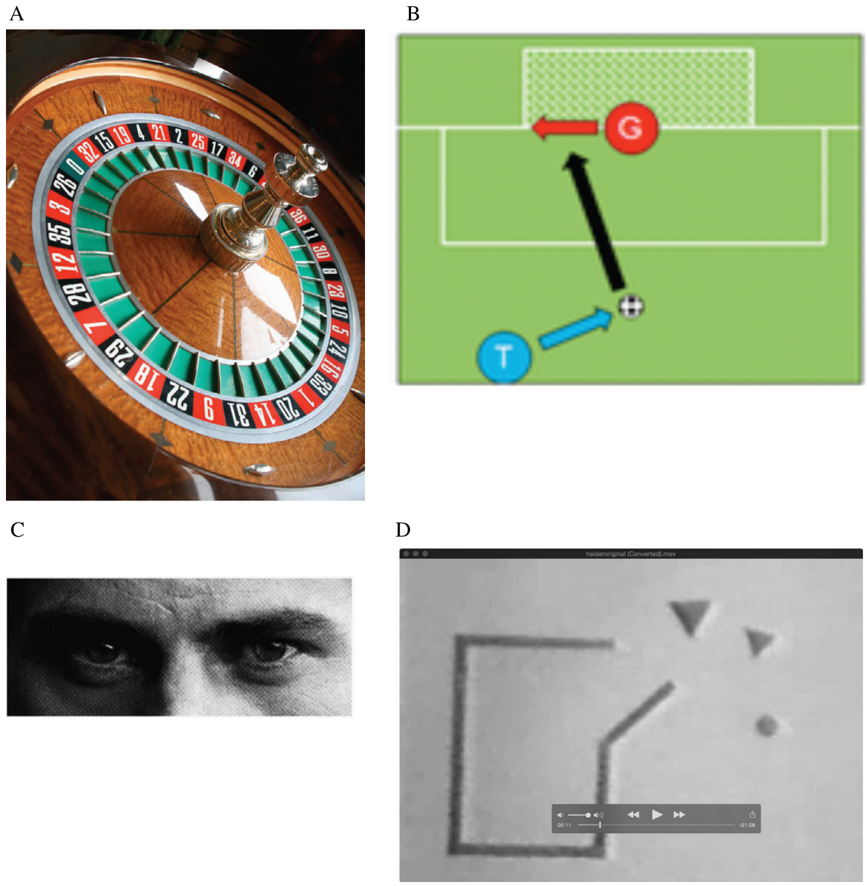

Fig. 1. (A) Traditionally, participation in financial markets has been treated as analogous to playing games with a physical chance device such as roulette, shown here. (B) Soccer (football) penalty kicks can be analysed as a matching-pennies game where the kicker (''T'') attempts to mis-match (chooses a corner different from the one the goal keeper chooses) while the goal keeper (''G'') attempts to match (matches the corner). Key to playing this game is understanding the intentions of one's opponent, and hence, applying ''Theory of Mind'' (ToM). (C) In the eye gaze test, the subject is asked to characterise the intentions reflected in the gaze. The test is a standard way to score social cognition. (D) In the Heider movie, two triangles and a circle move randomly across the screen. The moves can be predicted successfully by attributing beliefs and intentions to the geometric objects.

An intentional stance is ''the strategy of interpreting the behaviour of an entity treating it as if it were a rational agent who governed its choice of actions by a consideration of its beliefs and desires'' (Dennett, 2009). It contrasts with a physical stance, which aims at predicting the behaviour of an entity through analyses of its physical (if innate objects) or bio-physical (if living organisms) make-up. It also contrasts with a design stance, which is the ability to predict from an understanding of the functionality of an object or organism (what is it meant to accomplish?).

One could view the intentional stance as a crucial step towards scientific understanding. Overwhelmed with complexity, humans at first attempt an intentional stance when predicting outcomes of a system (physical, biological, social). One could interpret attribution to ''higher'' (divine) powers as an example of this stance (which incidentally are often modelled in the image of the human himself; cf. Greek mythology). As understanding increases, they switch to a design stance — they predict behaviour from its purpose. When eventually the (bio-)physics becomes understood, humans take a physical stance.

An intentional stance can only be defended if it works. To use an analogy due to Dennett (2009), one could envisage an ''astrological stance'', whereby one attempts to predict an entity based on the alignment of stars associated with it, and indeed many people use this to predict their own and others' future. To our knowledge, it has yet to be demonstrated to work. In contrast, we will provide evidence here that shows that an intentional stance does work in a financial risks context, albeit not always.

We humans often use the intentional stance to better understand complex systems. Outcomes of a blind bio-physical process called ''evolution'' speak better to one's imagination if cast in terms of an intentional system (which it is NOT!). Consider, for instance, the following phrase: ''When evolution discovers regularity or constancy in the environment, it designs adaptations that tacitly presuppose that regularity'' (Dennett, 1989). The phrase describes the outcome of evolution, but the verbs ''discover'', ''design'', ''presuppose'' imply intentionality, which evolution does not possess.

Here we explore, (i) to what extent the intentional stance is used in a financial markets context, (ii) whether and when it can be successful, and (iii) how variations in its use over time and across individuals explain choices and performance.

Traditional asset pricing theory does not treat financial markets as intentional. The assumption, in a static context, of perfect competition (prices are taken as given), and in a dynamic setting, of rational expectations (the future evolution of prices is taken as given; Radner, 1972), implies that investors are to merely optimise in the face of a system that generates outcomes ''as if'' it were a physical device beyond their control. When investors are given the equilibrium mapping from states to prices, as in dynamic asset pricing theory, they are not to test its veracity even if their own perception of how the economy works may generate a different mapping (Bossaerts, 1998).

And yet, finance scholars themselves regularly resort to the intentional stance. Indeed, prices in financial markets are explained in terms of a representative agent. It is true that a representative agent will exist from the moment Pareto efficiency is reached, but only in rare circumstances do the choices of this representative agent exhibit characteristics (e.g., preference parameters) that could be recognised as "human" (Constantinides, 1982). Discussions as to whether preference parameters estimated from historical financial markets data make "sense" tacitly assume that the representative agent optimises like a real human being; rarely is this admitted openly (Epstein et al., 2014).

The intentional stance requires what psychologists have been referring to as *Theory of Mind* (ToM). This is the capacity to (i) detect intentionality in one's environment, (ii) "read" the intentions, and (iii) act successfully upon them (Frith and Frith, 2005). To detect intentionality is pretty much engrained: even non-human primates pause and wonder when they observe a physical object that violates the laws of physics, as if primates knew Newtonian physics (Uller and Nichols, 2000). Infants generally cannot imagine that others may have different beliefs than they themselves have. The "chocolate in the drawer" problem tests the development of the capacity to separate one's own beliefs from that of another person (Gallagher and Frith, 2003). To successfully act on a reading of the beliefs and desires of others requires sophisticated social cognition.

Quite a bit is known about ToM, so we should digress and explain how to discern its presence, scope and quality. We do so in the next section. In Sections 3–5, we discuss evidence from three separate studies on financial decision-making. Section 6 discusses the implications.

2. Theory of Mind (ToM)

Psychologists have developed several tests to determine the extent to which a subject applies ToM. Here, we discuss two, suitably adapted to look like tests familiar from economics experiments, where subjects are paid for performance.

One is the eye gaze test (Baron-Cohen et al., 2001; Fig. 1(C)), where the subject is asked to study a picture of a person's eye gaze and to choose one among four possible adjectives which best reflects the beliefs or desires of that person. Here, ToM is real: there is consensus that the person whose eye gaze is depicted actually believed or desired as in the correct answer.

The second one we have used in the past is the Heider test. A film of moving geometric objects (two triangles of differing size; one circle) is shown (Heider and Marianne, 1944); see Fig. 1(D). Most people discern a pattern in the moves, namely, a situation where a third person (the circle) intercedes when one person (the small triangle) is being bullied by another person (the large triangle). Psychologists would ask subjects to describe the scene, and then measure the extent of application of ToM by counting the number of belief- or intention-related terms in the description. To determine whether this helps, we add forecasting to the task: every 5 s we stop the movie and ask the subject to forecast whether the two triangles will be farther apart or closer together 5 s later. Here, ToM does work: to imagine a bullying scene helps in forecasting the future distance between the triangles. There is no true intentionality in the moves of the objects, and as such, ToM works only "as if".

We have also looked at brain activation for neural evidence consistent with the hypothesis that ToM is being applied. One does so by means of, e.g., functional magnetic resonance imaging (fMRI) while a subject is watching a replay of financial markets. Two regions of the human brain tend to be particularly active in situations where subjects appear to be applying ToM, namely, the dorsomedial Prefrontal Cortex (abbreviated dmPFC, and sometimes referred to as the paracingulate cortex), and the Temporo-Parietal Junction (TPJ) — or more precisely the posterior part of the Superior Temporal Sulcus (STS), as well as the Inferior Parietal Lobule (IPL) (Fig. 2(A) and (B)).

It is not sufficient that these two regions (dmPFC; TPJ) activate in order to be sure that the subject is taking an intentional stance, because functionality of dmPFC and TPJ is not limited to social cognition. This "reverse inference" is to be avoided: because a brain region shows a particular functionality in one task, and activates in another task, this does not mean that the same functionality is at work in the second task (Poldrack, 2006). Among others, dmPFC is also engaged in tracking changes in latent driving variables behind stochastic outcomes, as in Kalman filtering. To be sure, Kalman filtering is not unlike ToM: the decision-maker uses the evolution of outcomes in order to infer the latent variables that caused the outcomes. In ToM, the latent variable is the mind of one's opponent; in Kalman filtering, it is a underlying physical state of the world. A clear neurobiological separation has yet to be found between inference about another person's mind and about an abstract underlying state (Suzuki et al., 2015).

fMRI is an imaging technique whereby the brain is exposed to a strong magnetic field while radio pulses are being applied. This allows the investigator to detect concentrations of oxygen-rich blood, a sign that locally the neural cluster has activated. The signal to come out of the analysis is referred to as the "BOLD" (Blood Oxygen Level Dependent) signal. The BOLD signal provides an indirect measure of neural activation. fMRI facilitates localisation.

To become more confident that ToM generated the activations, we go beyond localisation (which brain region activates?) and probe computations (what is the brain region computing?). We posit a computational model of ToM, the output of which changes with the continuously varying inputs. We require that the model successfully predicts the subject's actions. We then correlate key variables in the model (e.g., uncertainty; predictions of the opponent's reactions) with the fMRI signal. Beyond building confidence that ToM is at work, this *parametric* form of fMRI (O'Doherty et al., 2003) leads one to probe *which* theory (of mind) subjects are applying. In addition, behavioural heterogeneity in the application of ToM should be attributable to the relative strengths of the neural signals.

Parametric fMRI can best be illustrated by means of a non-controversial example of ToM. Consider the strategic game of matching pennies, played with a real (human) opponent. Take the case when the payoffs are asymmetric: the matcher's payoffs change depending on the action which she matches; see Fig. 2(C); there, we refer to it as the "inspection game", and call the matcher "employer", and the mismatcher "employee". The Nash equilibrium of this game requires both players to mix: they cannot always make the same choice because the opponent will exploit this and respond with the action that is best for her. In other words, a player should realise that he has an *influence* on the beliefs, and hence, intentions of the opponent. As a result, it is reasonable to conjecture that ToM is crucial in this type of game.

Not only did players indeed attempt to influence the beliefs of the opponent; brain activation in dmPFC correlated with the predicted effect of one's current choice onto future beliefs, and hence, future choices, of the opponent; activation in TPJ reflected the ensuing prediction error, when the opponent ended up acting differently from predicted. The activations are stronger for participants whose choices reveal more intense attempts at influencing; see Fig. 2(D) (Hampton et al., 2008).

Proof that the activation in TPJ is *causal* has recently emerged. In Hill et al. (2017), Transcranial Magnetic Stimulation (TMS) was applied to TPJ. TMS is a technique that allows the researcher to temporarily disrupt signals in a brain region. Effectively, TMS creates a temporary brain lesion. The authors showed how, in the same game as in Hampton et al. (2008), participants were no longer attempting to influence the beliefs of their opponent. Behaviour, as well as neural signals in the brain regions to which TPJ projects (such as dmPFC), changed.

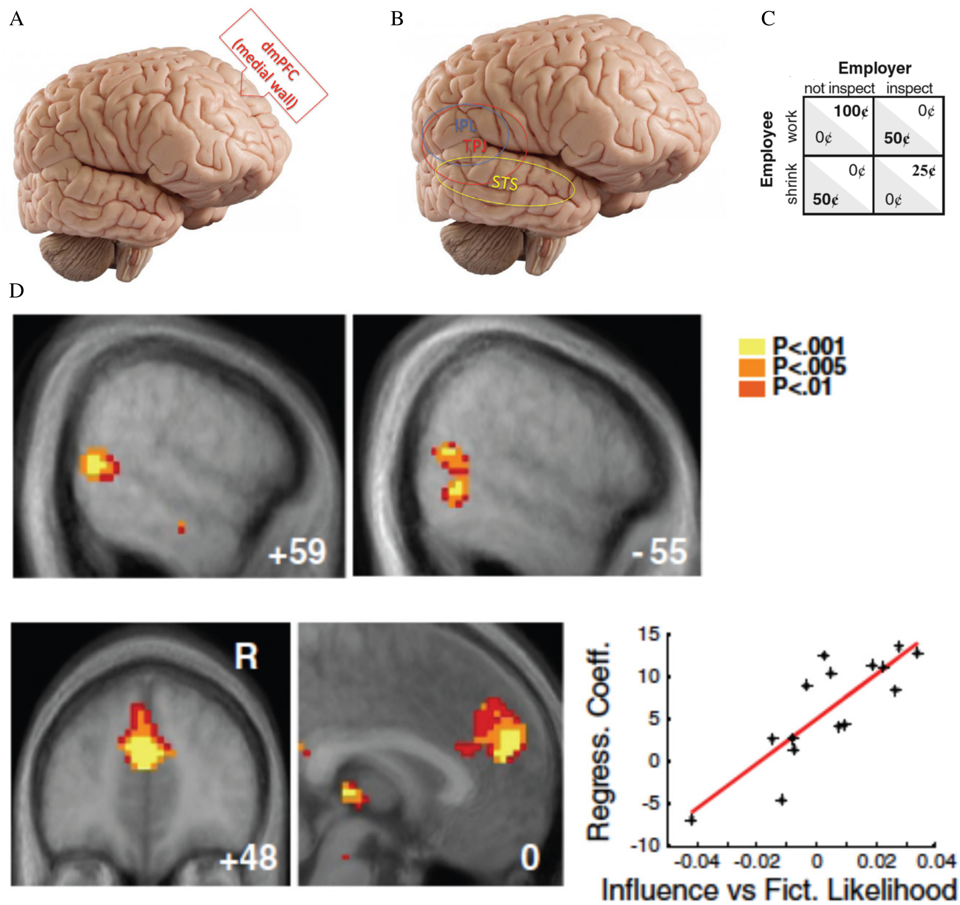

Fig. 2. Dorso-medial Prefrontal Cortex dmPFC (A) and Temporo-parietal Junction TPJ/Superior Temporal Sulcus STS/Inferior Parietal Lobule IPL (B) are activated when humans attempt to gauge the mind of an intentional agent (i.e., apply Theory of Mind). (C) Payoff matrix in an asymmetric matching pennies game called the inspection game, where an employee (row player) chooses to work or shirk when the employer (column player) has the option to inspect or not to inspect. (D) Activation in dmPFC reflects predictions of the perceived influence of one's actions on future actions of one's opponent (bottom), while activation in TPJ/IPL/STS reflects corresponding prediction errors (top). Plot: neural signals correlate more intensely with predictions when choices reveal stronger attempts to influence opponent beliefs as opposed to reacting to the opponent's past actions ("fictitious play").

ToM thus plays a crucial role in strategic games between two humans, and its neurobiological foundations are becoming understood. In the subsequent sections, we will see that ToM extends beyond the two-person setting, and provides insights into behaviour and neural activation also when a person is confronted with outcomes from larger-scale social interaction, such as transaction prices in financial markets.

3. Markets with insiders

In a first case where we conjectured that ToM could play a role in explaining subjects' financial decision-making, we replayed markets with insiders (Bruguier et al., 2010). The markets were part of an earlier controlled experiment where subjects could trade assets in an online continuous open-book exchange; more details are in Bossaerts et al. (2014). Assets paid a random liquidating dividend between 0 and 50 (U.S.) cents. In a number of replications, "insiders" were given precise information of the final dividend; in others, nobody was given privileged information and hence the best guess of the payoffs was 25 cents.

We used a simple graphical display (Fig. 3(A)) to replay the flow of bids, asks and trades while we scanned the observer's brains using fMRI. The observer was given the same information as uninformed traders in the original experiment. She was asked to monitor trades, and indicate with a key press when a trade had occurred. While participants rarely made mistakes, the task did require continuous attention, because orders arrived at the rate of one per 0.7 s, and transactions occurred once every 3.2 s. The observer was exposed to the risk in the market, as follows: at the beginning of a replication, she had to decide whether to take a long or short position in ten (10) units of the traded asset at a price of 0.25 dollar per unit. At the end of the replication, the subject earned the dividends on the 10 units minus the purchase price (if long) or the purchase price minus the dividends (if short).

We contrasted brain activation across replications with and without insiders. We focused on brain activation that increased as transaction prices moved away from 0.25 dollar, the best estimate of the asset's payoff absent privileged information. We expected neural activation in ToM regions when insiders were attempting to profit from buying if their information about final payoffs was favourable or selling if information was unfavourable, pushing the price above or below 25 cents respectively. Without insiders, price movements away from 0.25 dollar do not reveal anything about final payoffs; at best they reveal that supply is tight (price above 25 cents) or plentiful (prices below 25 cents). Besides anterior insula and amygdala, only dmPFC activated strongly, suggesting engagement of ToM (Fig. 3(B)). Anterior insula and amygdala are involved in tracking surprise (see Preuschoff et al., 2008; Prévost et al., 2011).

Consequently, we hypothesised that ToM skills would explain performance differences in a task where payment depended on the evolution of prices. We re-ran the experiment outside the scanner,

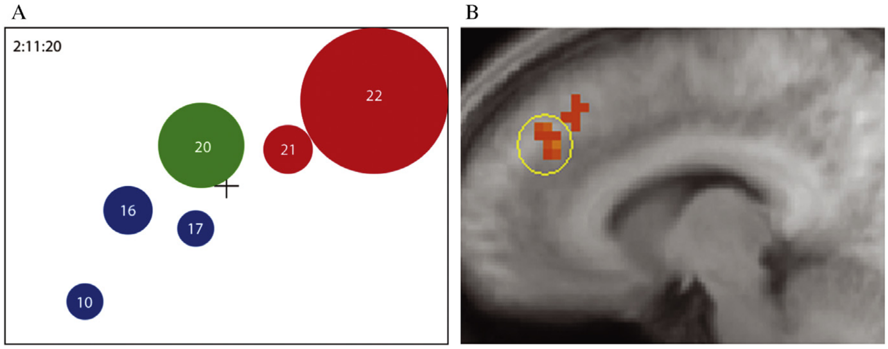

Fig. 3. (A) Graphical display of order flow. Red and blue bubbles depict asks and bids, respectively, at prices (in U.S. cents) written in the circle; size of the bubble indicates number of units offered (asks) or demanded (bids). Bubbles temporarily turn green when a transaction occurs. (B) fMRI signal shows higher brain activation in dmPFC when insiders are present than when there are no insiders. Shown is activation that correlates with the distance of the transaction price from value of the asset in the absence of information. The significance level (p) increases as colour becomes lighter, starting from red ($p = 0.001$).

but now stopped the market re-play every 5 s and asked subjects to guess whether the transaction price at the end of the subsequent 5 s interval would be greater than, equal to, or less than, the last recorded transaction price. We correlated performance on the financial markets prediction task with scores on the ToM Heider movie test and on the ToM eye gaze test (see above). The correlation between financial market forecasting performance and forecasting in the Heider movie equalled 0.35 ($p < .05$). Evidently, in financial markets, an intentional stance helps in forecasting prices based on order and transaction flow, just like it helps in forecasting movements of geometric objects when these reflect some kind of familiar social situation (the Heider movie test). Correlation with scores on the eye gaze test were lower: 0.30 ($p = 0.05$).

In the eye gaze test, ToM is not merely "as if:" eye gazes depict actual intentionality. This contrasts with the order and transaction flow in a markets setting, or with the movements of triangles and circle in the Heider test, where intentionality is only "as if". In a financial markets setting, order and trade arrivals are the result of interaction between (twenty) subjects; while each subject could be considered intentional, the result of their interactions is not. Note that the traditional theoretical concept used to analyse a market with insiders, the noisy rational expectations equilibrium (Grossman, 1977; Bossaerts et al., 2014) likewise assumes that agents consider markets as non-intentional: all agents posit the correct (noisy) mapping from states/signals to prices, and optimise against it; they do not question the mapping; they certainly do not attempt to influence it.

Curiously, we found no correlation between performance on the financial task and scores on a test of mathematics and logic skills typically used in the financial industry. Lack of correlation between performance and mathematics skills re-emerged in a later study on performance in a multi-person beauty contest game (Coricelli and Nagel, 2009). There too, ToM skills correlated significantly with performance in the game, while mathematics skills did not.

Participation in financial markets is complex. It requires not only the ability to predict outcomes, but also capacity to convert desired position changes into successful trades. Here, we have related ToM to the first facet only. The relevance of ToM to explain forecasting performance in markets with insiders has recently been confirmed in an independent study (Corgnet et al., 2017). However, in that study, ToM appears to correlate less with trading performance. That is, better ToM skills do not translate into better trading performance. Evidently, biases traditionally associated with non-intentional risks, such as gamblers'/hot hand fallacies, are better predictors of trading success.

Still, we should emphasise that we do not claim that the intentional stance, and hence, application of ToM, guarantees eventual trading success. We only demonstrate that ToM is relevant in explaining investor choices. Whether these ToM-influenced choices are beneficial is a different matter. As we shall see next, ToM is associated with riding financial bubbles that eventually crash. There, the intentional stance ultimately leads to low performance.

4. Bubbles and crashes

We replicated the above experiment, but instead of markets with insiders, we replayed a standard multi-period market experiment where an asset paid a stochastic dividend at regular points in time (fifteen in total) and expired worthless (De Martino et al., 2013). The value of the asset decreased each time a dividend was paid; the decrease equalled the expected period dividend. Bubbles (prices above the declining fundamental value) regularly emerge in this setting (Fig. 4(A)) though not always (Fig. 4(B); Smith et al., 1988). This is fortunate: we could contrast behaviour and brain activation in sessions that were identical except for the amount of mis-pricing. Importantly, subjects were informed that nobody had privileged ("insider") information about upcoming dividends.

Compared to the previous markets replay, we changed one aspect of the protocol: instead of fixing asset holdings at the beginning of a replication and asking for confirmation of transactions, we paused the replay at regular points in time, and allowed subjects to change positions at the best standing bid or ask (Fig. 4(C)). Consequently, the task was far more involved, and brain activation correspondingly complex.

Here too, we discovered increased activation in dmPFC during bubble sessions when contrasted with non-bubble sessions. Activation increased in the value of a subject's portfolio, which were inflated during a bubble, though only if the subject was "riding" the bubble (Fig. 4(D)). Importantly, the neural signal in dmPFC was significantly stronger for subjects with better ToM skills, but only during bubble sessions. We measured ToM skills by the score on the eye gaze test.

In an attempt to uncover the computations that dmPFC was engaged in during bubble sessions, we correlated a measure of irregularities in order arrivals with activation in dmPFC. Specifically, we applied a rolling-window statistic that measured non-homogeneities in the order arrival process (i.e., random changes in the arrival rate). We referred to it as the Poisson Inhomogeneity Detector (PID). In bubble sessions, activation in dmPFC indeed correlated with PID. Correlation increased with a subject's susceptibility to ride the bubble. We concluded that learning using ToM is based on erratic order arrivals. This is not unlike the way in which humans recognise intentionality in movements of physical objects: when such an object does not follow a straight path, humans attribute intentionality (Uller and Nichols, 2000).

A remark is in order. Our PID statistic is related to the PIN metric that has been claimed to track the presence of insiders

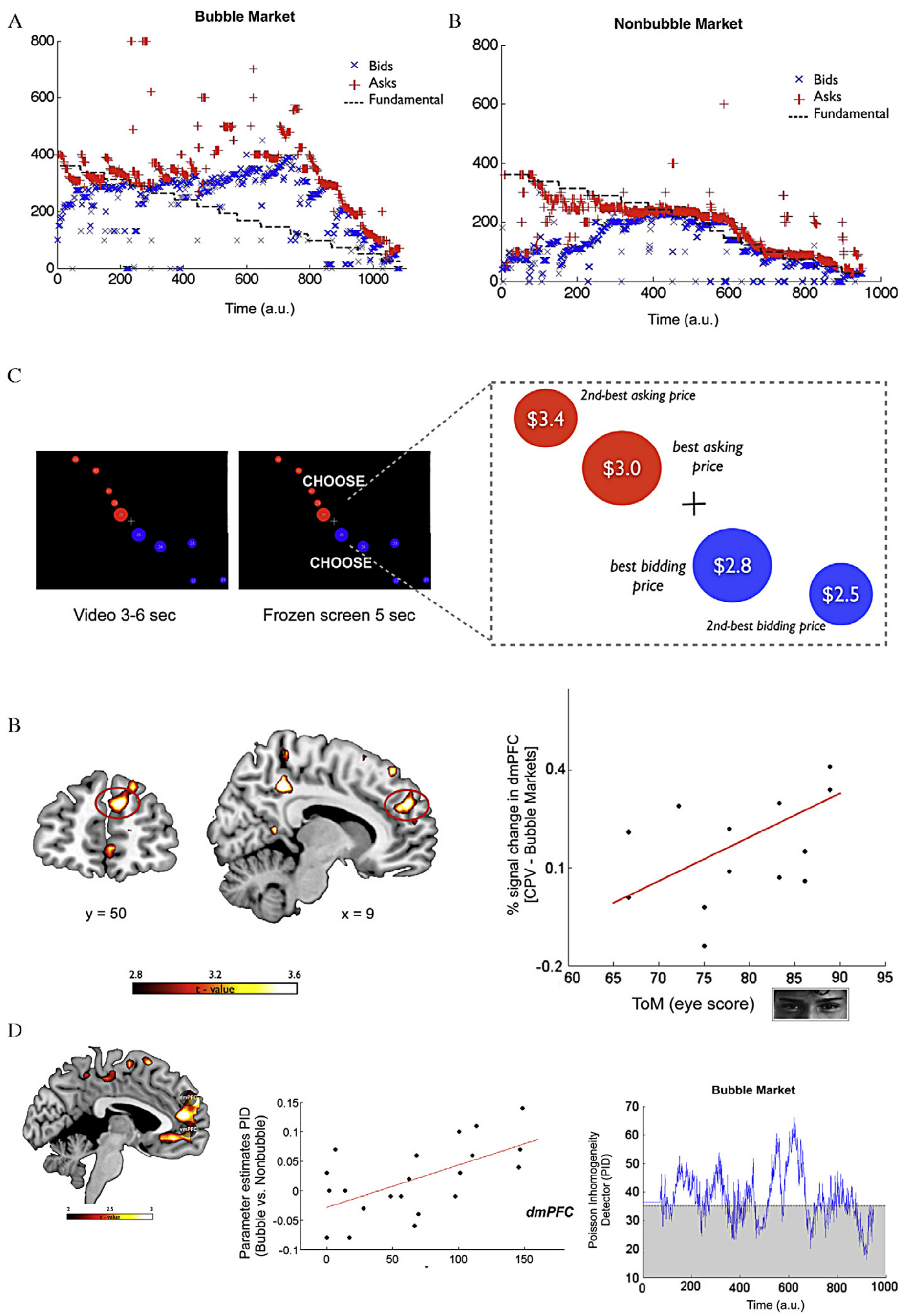

Fig. 4. (A) Typical bubble in a multiperiod experimental market session where the traded asset has a declining fundamental value (value declines periodically after dividends are paid). Blue: bids; Red: asks. Dashed line: fundamental value. (B) A bubble does not always appear. (C) During market replay in the scanner, the subject is asked periodically whether to change investment in the asset. (D) dmPFC activation increases in the current portfolio value in bubble sessions (contrast with non-bubble sessions); increase is stronger for subjects with higher ToM skills as measured by score on eye gaze test. E. In bubble sessions (only), activation in dmPFC correlates with a measure of inhomogeneity of the order arrival process (PID). Right: PID measure in one bubble session; measure is significant at $p = 0.05$ when it increases beyond the grey zone. Colours in B and D refer to level of t statistic as indicated below each brain mapping.

(Easley et al., 2002). Importantly, here there are no insiders. Consequently, PID and PIN metrics may not correctly identify whether there are insiders in the marketplace.

5. Financial contagion

Our last evidence concerns financial contagion. We studied to what extent investors' risk attitudes could be affected by observing risky choices of another agent (Suzuki et al., 2016). Here, the agent is intentional, either a real human being (though in reality it was a computer programme emulating human choices recorded earlier), or a computer programmed to emulate human choices (an "artificially intelligent" agent). In order to ensure subjects paid attention, we not only asked subjects to observe choices of the agent, but to predict their choices as well, after a period of observation (Fig. 5(A)). Risk attitudes revealed in the choices of the agent significantly affected the observer's own risk attitudes. Neuro-biologically, observing the agent's choices biased the observer's neural risk signals. These signals emerge in a sub-cortical structure crucial for instrumental learning, namely, the caudate (Fig. 5(B)). Significantly, neural signals correlating with expected reward or probability of reward were unaffected. That is, the

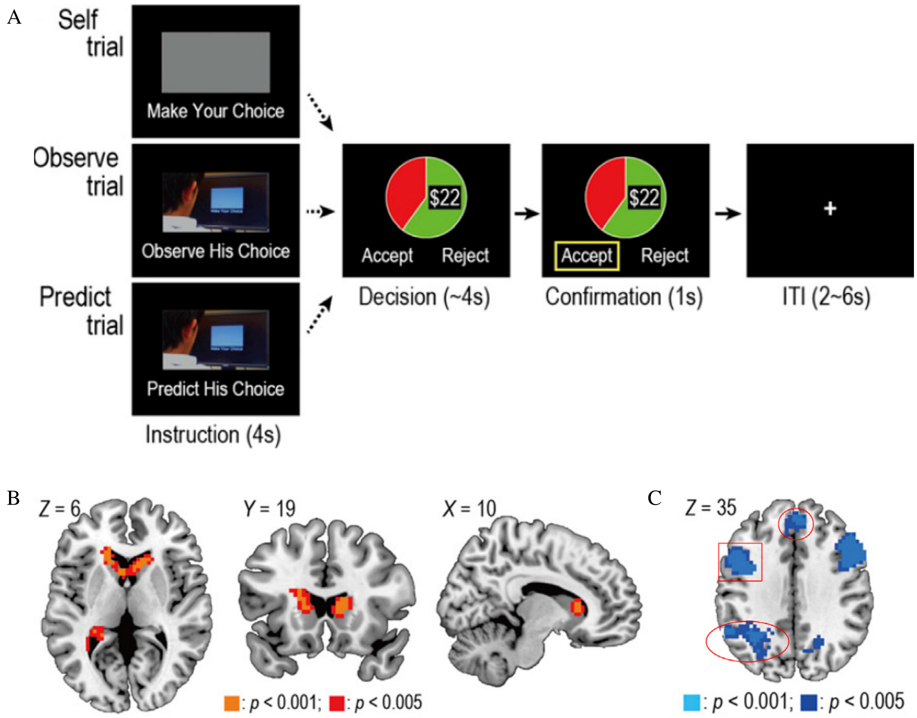

Fig. 5. (A) Blocks contained "Self", "Observe", and "Predict" trials, where subjects had to choose whether to accept a gamble (Self), observe another agent make a choice (Observe) or predict the other agent's choice (Predict). (B) Risk-related activation in caudate became biased through observing and predicting the other agent's choices, thus generating financial contagion. (C) Extent of belief updating correlated with activation in ToM regions (dmPFC:circle; TPJ: oval), as well as a region involved in tracking Bayesian surprise (dlPFC: square).

observer's brain activation did not reveal any tendency towards optimism/pessimism; instead, risk assessment changed.

Activation in ToM regions again emerged (Fig. 5(C)). This is perhaps not surprising because subjects were asked to observe the agent's choices and predict them. Here, dorsolateral Prefrontal Cortex (dlPFC) also activated. This region is known to generate neural signals of Bayesian surprise. Here, surprise concerns the extent to which the observer mis-predicted the agent's choices.

Note however that ability to predict did not correlate with extent of financial contagion (Fig. 5(D)). There is a mechanistic explanation for this: the link between prediction and contagion depended not only on ToM, but also on functional connectivity between dlPFC and caudate. This connectivity has been associated with executive function. As a result, contagion requires *both* ToM *and* good executive function.

The contagion effect from observing other participants' portfolios has recently been confirmed in a markets experiment. Specifically, diversification tends to increase when observing average portfolios of others, while diversification is reduced when broadcasting the (under-diversified) portfolios of recent winners. See Baghestanian et al. (2015).

6. Discussion

Evidence has emerged that humans treat risks associated with participation in financial markets as intentional, as if generated by a mindful agent with own intentions and desires. This contrasts with traditional and behavioural finance, where investors are assumed to view financial risks as non-intentional, as if generated by a physical device like a roulette wheel. The intentional stance at times helps agents – as in the case of markets with insiders – but at other times misleads them – as when deciding to "ride" a financial bubble. Not all investors have an equal inclination to take an intentional stance, and not all investors have equal capacity to apply Theory of Mind (ToM), and this leads to cross-sectional differences in behaviour and neural activation.

It is important to keep in mind that our claim is *not* that ToM is beneficial. Our claim is that the intentional stance is being used in financial decision-making, and that choice heterogeneity can be explained by differences in inclination to, and skill in, applying ToM.

It is equally important to appreciate that ToM is only part of the story. In the case of bubbles and financial contagion, conversion of plans into successful trades, i.e., execution, may blur the effects of ToM. We conjecture that ToM is relevant mostly in the investments sphere, where agents decide on portfolios on the basis of their predictions of market evolution. In contrast, trading engages a different layer of cognition, namely, executive control, and the two sets of skills (ToM; executive control) do not necessarily correlate. Corgnet et al. (2017) provides preliminary evidence that, in the context of markets with insiders, ToM skills do indeed help subjects predict prices, but ToM skills do not directly correlate with trading performance. Further evidence in favour of the hypothesis that ToM skills are necessary but not sufficient for successful trading is presented in Hefti et al. (2016).

One additional possibility is that the capacity for ToM interacts with other variables such as quantitative skills, so that one or the other is not sufficient for good performance; both are necessary. Alternatively, it might be that there is a non-linear relationship, so that some level of ToM ability results in a capacity to be vulnerable to, say, bubbles in financial markets, but those with higher ability are aware of the bubble and are better able to "get out" at the right time. Higher ability may be related to a trader's depth of reasoning (strategic sophistication; Coricelli and Nagel, 2009).

Recent neuroscience studies suggest however, that is not quantitative skills, but "gut feeling", which is missing in the theory. In the field, profits as well as tenure of professional traders increases with their ability to sense heartbeat (Kandasamy et al., 2016). The finding may not sound credible were it not that there is quite a bit of direct and indirect evidence for the importance of emotional reactions in a financial setting. Lo and Repin (2002) found that heartbeat of professional traders reliably tracks market volatility. Critchley et al. (2004) discovered that the ability to sense one's heartbeat increased with the size of a crucial brain structure, anterior insula. Preuschoff et al. (2008) showed that neural signals in anterior insula correlate with risk and with risk prediction errors. As such, the path from heartbeat sensation to trading success is traceable, and it involves anterior insula.

Remarkably, Smith et al. (2014) demonstrated that it is possible to use activation in anterior insula to predict which traders will successfully "ride a bubble". Using the same bubbles-and-crashes markets setting discussed in Section 4, they show how elevated anterior insula activation precedes timely exit from the bubble. The sub-region where activation predicts success is the same as the one where Preuschoff et al. (2008) detected risk prediction error signals. So, while Theory of Mind predicts who will ride a bubble in the first place, volatility-related "gut feeling" appears to determine who gets out in time.

Theoretical modelling of bubbles tends to treat bubble-riders as passive "trend-followers" (e.g., Hommes et al., 2008; Barberis et al., 2016). The research discussed here suggests that (some? all?) bubble-riders are actually anything but passive. They engage in forward-looking Theory of Mind, assessing the intentions of the market and attempting to profit from being one step ahead of the market. Bubble-riders behave more like the rational players in Brunnermeier and Morgan's "clock games" (Brunnermeier and Morgan, 2010).

Behavioural finance has been inspired by the heuristics program of Gigerenzer and Selten (2002), and the heuristics and biases program of Tversky and Kahneman (1974). These heuristics and biases concern non-intentional risk and uncertainty. We would advocate expanding the study of financial decision-making to include intentional stance. The argument is that investors often take an intentional stance in order to better comprehend complex systems such as financial markets. Whether they benefit from it depends on the setting, and, as we argued, "gut feeling", plus adequate executive control.

The two aspects of human cognition, heuristics and biases, and Theory of Mind, may not be orthogonal. For instance, one can envisage that gambler's fallacy emerges because of intentional stance. Humans may have experienced that, in strategic games, their opponents are generally not very good at "mixing": opponents create pseudo-random number sequences with runs that are too short (except under explicit instruction; see Rapoport and Budescu, 1992). As such, when humans perceive financial markets to be human-like, they expect reversals to happen sooner than when generated by a non-intentional risk source. The reversal expectation causes gambler's fallacy (Rabin and Vayanos, 2010).

Acknowledgements

The authors thank Carsten Murawski for several discussions about philosophy of mind and its relation to economic and scientific thinking, which helped shape this review; Carsten Murawski and Christo Karunanathan for comments on an earlier draft; and participants during presentations at the 2015 Auckland Finance Meetings (New Zealand), the 2016 Conference on Experimental Finance at Chapman University, the 2016 Australasian Meetings of The Econometric Society, the 2016 Tucson Meetings of the Society for Experimental Finance, at the University of Geneva, the University of New South Wales and at Macquarie University.

References

Baghestanian, Sascha, Gortner, Paul J., Van der Weele, Joel J., 2015. Peer effects and risk sharing in experimental asset markets.

Barberis, Nicholas, Greenwood, Robin, Jin, Lawrence, Shleifer, Andrei, 2016. Extrap olation Bubbles, Technical Report. National Bureau of Economic Research.

Baron-Cohen, Simon, Wheelwright, Sally, Hill, Jacqueline, Raste, Yogini, Plumb, Ian, 2001. The "reading the mind in the eyes" test revised version: A study with normal adults, and adults with asperger syndrome or high-functioning autism. J. Child Psychol. Psychiatry 42 (2), 241–251.

Bossaerts, Peter, 1998. Rational Expectations Equilibrium When Priors are Inconsistent. California Institute of Technology.

Bossaerts, Peter, 2009. What decision neuroscience teaches us about financial decision making. Ann. Rev. Financ. Econ. 1 (1), 383–404. http://dx.doi.org/10.1146/annurev.financial.102708.141514. http://www.annualreviews.org/doi/abs/10.1146/annurev.financial.102708.141514.

Bossaerts, Peter, Frydman, Cary, Ledyard, John, 2014. The speed of information revelation and eventual price quality in markets with insiders: Comparing two theories. Rev. Financ. 18 (1), 1–22. http://dx.doi.org/10.1093/rof/rfs049. http://rof.oxfordjournals.org/content/18/1/1.abstract.

Bruguier, Antoine J., Quartz, Steven R., Bossaerts, Peter, 2010. Exploring the nature of trader intuition? J. Finance 65 (5), 1703–1723.

Brunnermeier, Markus K., Morgan, John, 2010. Clock games: Theory and experiments. Games Econom. Behav. 68 (2), 532–550.

Constantinides, George M., 1982. Intertemporal asset pricing with heterogeneous consumers and without demand aggregation. J. Bus. 55 (2), 253–267; http://www.jstor.org/stable/2352702 (ISSN: 00219398, 15375374).

Corgnet, Brice, DeSantis, Mark, Porter, David, 2017. What makes a good trader? on the role of quant skills, behavioral biases and intuition on trader performance. J. Finance. Available at: SSRN: https://ssrn.com/abstract=3006695.

Coricelli, G., Nagel, R., 2009. Neural correlates of depth of strategic reasoning in medial prefrontal cortex. Proc. Natl. Acad. Sci. 106, 9163–9168.

Critchley, Hugo D., Wiens, Stefan, Rotshtein, Pia, Öhman, Arne, Dolan, Raymond J., 2004. Neural systems supporting interoceptive awareness. Nature Neurosci. 7 (2), 189–195.

De Martino, Benedetto, O'Doherty, John P., Ray, Debajyoti, Bossaerts, Peter, Camerer, Colin, 2013. In the mind of the market: Theory of mind biases value computation during financial bubbles. Neuron 79 (6), 1222–1231.

Dennett, Danie Clement, 1989. The Intentional Stance. MIT Press.

Dennett, Daniel C., 2009. Intentional systems theory.

Easley, David, Hvidkjaer, Soeren, O'Hara, Maureen, 2002. Is information risk a determinant of asset returns? J. Finance (ISSN: 1540-6261) 57 (5), 2185–2221. http://dx.doi.org/10.1111/1540-6261.00493.

Epstein, Larry G., Farhi, Emmanuel, Strzalecki, Tomasz, 2014. How much would you pay to resolve long-run risk? Am. Econ. Rev. 104, 2680–2697.

Frith, Chris, Frith, Uta, 2005. Theory of mind. Curr. Biol. 15 (17), R644–R645.

Gallagher, Helen L., Frith, Christopher D., 2003. Functional imaging of 'theory of mind'. Trends Cogn. Sci. 7 (2), 77–83.

Gigerenzer, Gerd, Selten, Reinhard, 2002. Bounded Rationality: the Adaptive Toolbox. MIT Press.

Grossman, Sanford J., 1977. The existence of futures markets, noisy rational expectations and informational externalities. Rev. Econom. Stud. 431–449.

Hampton, A.N., Bossaerts, P., O'Doherty, J.P., 2008. Neural correlates of mentalizing-related computations during strategic interactions in humans. Proc. Natl. Acad. Sci. 105 (18), 6741.

Hefti, Andreas, Heinke, Steve, Schneider, Frédéric, 2016. Mental capabilities, trading styles, and asset market bubbles: Theory and Exp.

Heider, Fritz, Marianne, Simmel, 1944. An experimental study of apparent behavior. Am. J. Psychol. (ISSN: 00029556) 57 (2), 243–259. http://www.jstor.org/stable/1416950.

Hill, Christopher A., Suzuki, Shinsuke, Polania, Rafael, Moisa, Marius, O'Doherty, John P., Ruff, Christian C., 2017. A causal account of the brain network computations underlying strategic social behavior. Nature Neurosci. 20 (8), 1142–1149.

Hirshleifer, David, 2001. Investor psychology and asset pricing. J. Finance 56 (4), 1533–1597; http://www.jstor.org/stable/2697808 (ISSN: 00221082, 15406261).

Hommes, Cars, Sonnemans, Joep, Tuinstra, Jan, Van de Velden, Henk, 2008. Expectations and bubbles in asset pricing experiments. J. Econ. Behav. Organ. 67 (1), 116–133.

Kandasamy, Narayanan, Garfinkel, Sarah N., Page, Lionel, Hardy, Ben, Critchley, Hugo D., Gurnell, Mark, Coates, John M., 2016. Interoceptive ability predicts survival on a London trading floor. Sci. Rep. 6.

Lo, Andrew W., Repin, Dmitry V., 2002. The psychophysiology of real-time financial risk processing. J. Cogn. Neurosci. 14 (3), 323–339.

Markowitz, H., Todd, G.P., Sharpe, W.F., 2000. Mean–Variance Analysis in Portfolio Choice and Capital Markets, Vol. 66. Wiley.

O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H., Dolan, R.J., 2003. Temporal difference models and reward-related learning in the human brain. Neuron 38 (2), 329–337.

Poldrack, Russell A., 2006. Can cognitive processes be inferred from neuroimaging data? Trends Cogn. Sci. (ISSN: 1364-6613) 10 (2), 59–63. http://dx.doi.org/10.1016/j.tics.2005.12.004. http://www.sciencedirect.com/science/article/pii/S1364661305003360.

Preuschoff, K., Quartz, S.R., Bossaerts, P., 2008. Human insula activation reflects risk prediction errors as well as risk. J. Neurosci. 28, 2745–2752.

Prévost, Charlotte, McCabe, Jonathan A., Jessup, Ryan K., Bossaerts, Peter, O'Doherty, John P., 2011. Differentiable contributions of human amygdalar subregions in the computations underlying reward and avoidance learning. Eur. J. Neurosci. 34 (1), 134–145.

Rabin, Matthew, Vayanos, Dimitri, 2010. The gambler's and hot-hand fallacies: Theory and applications. Rev. Econom. Stud. 77 (2), 730–778.

Radner, Roy, 1972. Existence of equilibrium of plans, prices, and price expectations in a sequence of markets. Econometrica 289–303.

Rapoport, Amnon, Budescu, David V., 1992. Generation of random series in two-person strictly competitive games. J. Exp. Psychol. Gen. 121 (3), 352.

Sen, Amartya, Pattanaik, Prasanta K., 1969. Necessary and sufficient conditions for rational choice under majority decision. J. Econom. Theory (ISSN: 0022-0531) 1 (2), 178–202. http://dx.doi.org/10.1016/0022-0531(69)90020-9. http://www.sciencedirect.com/science/article/pii/0022053169900209.

Smith, Alec, Lohrenz, Terry, King, Justin, Read Montague, P., Camerer, Colin F., 2014. Irrational exuberance and neural crash warning signals during endogenous experimental market bubbles. Proc. Natl. Acad. Sci. 111 (29), 10503–10508.

Smith, Vernon L., Suchanek, Gerry L., Williams, Arlington W., 1988. Bubbles, crashes, and endogenous expectations in experimental spot asset markets. Econometrica (ISSN: 00129682) 56 (5), 1119–1151. http://www.jstor.org/stable/1911361.

Suzuki, Shinsuke, Adachi, Ryo, Dunne, Simon, Bossaerts, Peter, O'Doherty, John P., 2015. Neural mechanisms underlying human consensus decision-making. Neuron 86 (2), 591–602.

Suzuki, Shinsuke, Jensen, Emily L.S., Bossaerts, Peter, O'Doherty, John P., 2016. Behavioral contagion during learning about another agent's risk-preferences acts on the neural representation of decision-risk. Proc. Natl. Acad. Sci. 113 (14), 3755–3760.

Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. Science 185, 1124–1131.

Uller, Claudia, Nichols, Shaun, 2000. Goal attribution in chimpanzees. Cognition 76 (2), B27–B34.