The Efficient Markets Hypothesis Does Not Hold When Securities Valuation is Computationally Intractable

By Shireen Tang and Harvey Huang and Elizabeth Bowman and Nitin Yadav and Carsten Murawski and Peter Bossaerts*

Draft: January 18, 2017

We study a situation where all market participants are endowed with the same information ex ante but valuation of securities is computationally intractable. More specifically, security values are given by the solution of a 0-1 knapsack problem, an NP-hard computational problem. We demonstrate experimentally that the quality of security prices decreases substantially as the computational complexity—that is, the amount of computational resources required to solve the knapsack problem—increases. We also show that market participants who were more successful at solving the knapsack problem, either due to skill or chance, made more profit. At the same time, market prices did not reflect security values, even though some participants knew them. These results suggest that the Efficient Market Hypothesis does not hold when computational complexity of security valuation is high. They also suggest that, given the high computational complexity of many economic problems, the Efficient Market Hypothesis is only plausible under very specific, and possibly rare, circumstances.

Keywords: Efficient Markets Hypothesis, Computational Complexity, Financial Markets, Grossman-Stiglitz Paradox, Hirshleifer Effect, Intellectual Discovery, Patents, Prediction Markets

* Brain, Mind & Markets Laboratory, Department of Finance, The University of Melbourne, Victoria
The Efficient Markets Hypothesis (EMH) is one of the foundations of modern asset pricing theory. Markets are called “efficient” when “security prices fully reflect all available information” [Fama (1991), p. 1575]. It is usually studied in a setting where the value of securities is a function of information that is distributed among market participants, and correct valuation requires aggregation of the different information sets (Grossman and Stiglitz, 1976). It is assumed, usually implicitly, that aggregation of information is feasible, in particular, that all computations necessary to aggregate the information, in the form of an expectation, for example, is possible. In the rational expectations equilibrium of such an economy (the “Radner” perfect foresight equilibrium; Radner (1972)), prices perfectly reveal values, and hence, “security prices fully reflect all available information.” Starting with Plott and Sunder (1982, 1988), controlled experiments have confirmed this prediction in the case where security payoffs are pure common-value, or, if not, private payoff schedules are known, and markets are organised as centralised double auctions.

It has been pointed out previously that EMH does not always hold though. If information is costly to obtain, then no agent has any incentive to acquire information; everyone should just wait for prices to settle, at which point information could readily be obtained from prices. Informed agents, that is, those agents who acquired information, would not be able to recuperate the cost of their information acquisition. This is one case where EMH does not hold (Grossman and Stiglitz, 1976). Sunder (1992) shows that this “Grossman–Stiglitz Paradox” (GS Paradox) is not only of theoretical concern. He verified the prediction experimentally: participants do not acquire information if information comes at a fixed cost,

---

1. The market microstructure literature sometimes takes a different approach, where better informed compete on the basis of the same information bit. See, e.g., Holden and Subrahmanyam (1992). There, the issue is whether prices eventually reveal that information bit, and how. See Bossaerts, Frydman and Ledyard (2014) for an experimental investigation.

2. The ability of decentralised markets to aggregate dispersed information is hotly disputed, with Wolinsky (1990) and Duffie and Manso (2007) taking opposite views. See Asparouhova and Bossaerts (2016) for experimental evidence.
and that the price of information decreases to zero if information is auctioned.

All analyses of EMH to date have focused on situations where the information necessary to value securities correctly, is either endowed or can be acquired, and values of securities can easily be computed based on this information. There are many situations where the latter assumption may not hold, that is, where valuation requires processing of information that requires a large amount of computational resources, possibly more than are available in the Universe. The amount of computational resources required to solve a given computational problem is referred to as its “computational complexity”\(^3\), and problems that require resources beyond those available are referred to as “computationally intractable” (Cook, 1983).\(^4\) Here, we study a situation where valuation of securities is computationally intractable.

More precisely, we investigate a setting where the value of securities is given by the solution of a 0-1 knapsack problem (KP), a combinatorial optimisation problem [REF]. In the KP, the decision-maker is asked to find the sub-set of items of different values and weights that maximises the total value of the knapsack, subject to a weight constraint. The KP belongs to the complexity class NP-hard [REF]. This means that there is no known algorithm that finds the solution and is efficient, that is, can compute the solution in polynomial time.\(^5\) Intuitively, it is ‘easy’ to verify that a particular set of items achieves a given total value, but

---

\(^3\)We use the term “complexity” as defined by computer science (Arora et al., 2010). Recent analyses of complexity in finance and economics either take a less formal approach (e.g., Skreta and Veldkamp (2009)), or model complexity in probabilistic terms (e.g., Spiegler (2016)).

\(^4\)Computer scientists would characterise the standard EMH setting as one of of finite sample complexity (Valiant, 1984) and low computational complexity. That is, correct valuation of securities only requires a finite number of samples from the distribution of payoffs, and computing security values is “efficient”.

\(^5\)In complexity theory, an algorithm is called “efficient” if the rate at which the amount of time to compute the solution grows as the size of the computational problem increases, is upper-bounded by a polynomial. The class of computational problems whose solutions can be computed efficiently is called P (for polynomial time). The class of problems for which no efficient algorithms are known, but whose solutions can be verified in polynomial time, is called NP (for non-deterministic polynomial-time). The KP can be solved with dynamic programming. This algorithm is polynomial in the number of items; however, the memory required to implement dynamic programming grows exponentially. Therefore, dynamic programming substitutes exponentially growing time to address memory for exponentially growing time to compute. Experiments demonstrate that humans do not appear to use dynamic programming when solving the KP, which is not surprising: human working memory is very small (Murawski and Bossaerts, 2016).
it is hard to find the set of items with the highest total value.

The KP permeates economic life, from low-level cognition (optimal inattention; Sims (2006)) to high-level cognition (portfolio optimisation; Kellerer, Pferschy and Pisinger (2004)). We therefore argue that it provides a realistic setting to test whether EMH holds in the presence of high computational complexity. Like computers, humans struggle to find solutions to instances of the KP; they differ in their ability to find solutions; and they display substantial heterogeneity in solution approaches (Murawski and Bossaerts, 2016). Therefore, the KP naturally generates a situation of information heterogeneity as a consequence of computational complexity.

Our experiment was organised as follows. We endowed participants with shares of several securities, each of which corresponded to an item in a given instance of the KP. All securities lived for a single period, after which they paid a liquidating dividend. The dividend equalled one dollar if the corresponding item was in the solution of the instance of the KP; and zero otherwise. After markets opened, participants traded the securities in a computerised continuous open-book system (a version of the continuous double auction where infra-marginal orders are kept in the system until cancelled). All participants were provided with the same information about the instance to be solved, that is, each items’ value and weight, and the total capacity of the knapsack. They had access to a computer program where they could try out candidate solutions.

In the following, we report results that address the following questions. (i) Do security prices reveal the solution of the KP instance? That is, were securities corresponding to items in the optimal knapsack priced at $1, and others at $0? (ii) Did informed traders, defined as participants who found the solution of the instance, make money? (iii) Do uninformed traders, defined as participants who

---

6 We used the software Flex-E-Markets (http://www.flexemarkets.com).
7 The program is part of the ULEEF GAMES suite and can be accessed at http://uleef.business.utah.edu/games.
8 It should be pointed out that, while some participants at times may found the optimal solution, they might not have realised that it was indeed the optimal solution Murawski and Bossaerts (2016).
did not find the solution, attempt to read information from prices? In particular, were uninformed traders able to improve their attempts as a result of events in the marketplace?

The remainder of the paper is organised as follows. In the next section, we briefly summarise the experimental paradigm. We then present the results. In the third section, we discuss the implications of our findings, in particular, how they related to the Grossman-Stiglitz Paradox and the Hirshleifer Effect; why past tests of EMH on historical field data may need to be reinterpreted; our interpretation of markets as a mechanism to incentivise intellectual discovery; for the promise of prediction markets; and for the Church-Turing thesis in the theory of computation [REF].

I. Experimental Design

Participants

Participants were recruited from The University of Melbourne community, in four experimental sessions with 18 (one session) or 20 (three sessions) participants per session. To be eligible, participants had to be current students of The University of Melbourne aged between 18 to 30 years with normal or corrected-to-normal vision. The final sample included a total of 78 participants (age range: 18 to 26 years, mean age: 22, standard deviation: 4, gender: 44 male, 34 female). The study was approved by The University of Melbourne Human Research Ethics Committee (Ethics ID: 1647059.1), in accordance with the World Medical Association Declaration of Helsinki. All participants provided written informed consent.

Task

Participants attempted five instances of the 0-1 knapsack problem, while simultaneously trading in an online marketplace.

This, of course, is a consequence of the nature of the KP.
Instances of the knapsack problem For each instance, participants selected a subset from a set of items with given values and weights, to maximise the total combined value subject to a total weight constraint. Instances were taken from two prior studies (Meloso, Copic and Bossaerts (2009); Murawski and Bossaerts (2016)). Formally, participants were asked to solve the following maximisation problem:

$$\max_{x_i} \sum_{i=1}^{I} x_i v_i \text{ s.t. } \sum_{i=1}^{I} x_i w_i \leq C \text{ and } x_i \in \{0, 1\},$$

where $i$, $w$, $v$ and $C$ denote the item number, item weight, item value and knapsack capacity, respectively. The number of items varied between 10 and 12. Instance parameters for the five instances are provided in the SOM. Detailed analyses of the instances can be found in the SOM of Murawski and Bossaerts (2016).

Knapsack interface: The KP instances were made available electronically on a computer interface where participants could try out different solutions. The software recorded every move of an item into and out of the knapsack.

Online marketplace: Each item in an instance mapped into a security in the online marketplace. Therefore, between 10 to 12 markets were available in each instance. Exchange was organised using the continuous double-sided open book system, like most electronic stock markets globally. Trading was done on the online experimental markets platform Flex-E-Markets. Participants traded for 15 minutes, or in later rounds, less.

The interfaces of the two systems, knapsack interface and online marketplace, are shown in See Fig. 1.

Participant instructions including the time line of a typical experimental session can be found in the SOM.

We recorded price data of every order and trade in the marketplace, and synchronised timestamps with the KP interface, so we could relate events in the

---

9 The application is part of a game suite called ULEEF GAMES; it can be accessed at http://uleef.business.utah.edu/games.

marketplace to participant attempts at solving the underlying KP instance.\footnote{The marketplace server became unreachable during the second round in the final session, so we have no trade data for that round. This denial of service originated with the server service provider, and hence was beyond our control.}

\textbf{Earnings}

Participants took positions in the items they believed to be in the optimal knapsack by buying shares of the corresponding security. They could also sell shares corresponding to items they believed not to be in the solution; short sales were not allowed, however. In every instance, participants were endowed with $25 in cash holdings (Australian dollars, approximately eighteen U.S. dollars), and 12 random shares. The price range of a share was bounded between $0 and $1.

Final earnings consisted of: (i) liquidating dividends for the shares held at market close; (ii) any change in cash holdings between the beginning and end of trading. Earnings were cumulative across instances. Additionally, participants...
received a fixed reward ($2) for submitting a proposed solution through the KS instance rendering software, as well as a show-up fee of $5.

*Initial Allocations*

We designed initial allocations of securities to induce trade, by concentrating individual endowments in particular markets. While initial allocations were randomised, they were “fair” in the sense that all participants received the same number of shares in correct items across the five instances. Initial allocations were such that $36 in liquidating dividends was paid per participant on average. Participants were not told that they had “fair” initial allocations. We imposed fairness in the belief that earnings would suffer from the Hirshleifer Effect if EMH were to emerge (see Discussion for more details).

Although the concepts of “risk” and “risk aversion” have yet to be defined precisely in the context of computational complexity, we intuited that uncertainty about the solution to a KP instance would induce participants to diversify holdings across multiple securities. As a result, participants would trade not only because of perception of superior information. We eliminated aggregate risk by ensuring that there were an equal numbers of shares across securities. This was meant to avoid price distortions that could arise from differences in relative supplies.

Additionally, payment for submission of a solution through the KP interface was fixed and independent of whether the submission was correct. This made it impossible for participants to hedge between trading in the marketplace and submission of solutions through the KP interface.

*Informed and Uninformed Traders*

All participants were given the same information about the KP instance in a session. Thus, there was no information heterogeneity and no information asymmetry ex ante. However, we did not provide participants with the solutions,
and since our instances were hard and performance variable (the proportion of participants who submitted the correct solution varied between 6.4% and 60.3%; see SOM for details), information heterogeneity (and asymmetry) arose spontaneously as participants started to search for the correct solution. We define a participant who submitted the correct solution as an “informed trader,” and one who did not, as an “uninformed trader.”

II. Results

Descriptive Statistics

Each of the 78 participants solved five instances of the KP (390 attempts in total). We first looked at computational performance of participants, that is, participants’ ability to find the optimal solutions of instances. To this end, we examined the proportion of participants that were able to solve an instance. Overall, 37.2% of attempts were correct. Performance varied both by participant (min = 0, max = 1, SD = 0.26) and experimental session (min = 0.33, max = 0.47, SD = 0.06).

Next, we investigated whether computational performance in an instance was related to the instance’s complexity. We measured instance complexity with the Sahni-\(k\). This metric increases with both the number of computational steps and the amount of memory required to solve an instance. Intuitively, Sahni-\(k\) is equal to the number of items that have to be selected into the knapsack before the knapsack can be filled up using the greedy algorithm to find the solution. The greedy algorithm fills the knapsack by selecting items in decreasing order of the ratio of value over weight until none of the remaining items fits into the knapsack. If Sahni-\(k\) equals 0, the greedy algorithm generates the solution of the instance. If \(k\) is greater than 0, then Sahni algorithm generates all feasible sets of items of cardinality \(k\), fills up the knapsack using the greedy algorithm and finds the set

12As mentioned before, the nature of the KP is such that even informed traders may not have been aware that they found the solution.
of items with the highest total value. The Sahni-$k$ of the instances in this study varied from 0 to 4.\textsuperscript{13}

The proportion of participants who solved the instance correctly decreased from 60.3\% when Sahni-$k$ was equal to 0, to 6.4\% when Sahni-$k$ was equal to 4. To test the negative relation between computational performance and Sahni-$k$, we estimated a mixed-effects model with a binary variable set to 1 if an attempt was correct as dependent variable (0 otherwise), a fixed effect for Sahni-$k$ and random effects for participant and experimental session on the intercept. We found a significant main effect of Sahni-$k$ ($\beta = -0.578$, $p < 0.001$). The pattern confirms the validity of Sahni-$k$ as a measure of instance difficulty for humans, first documented in Meloso, Copic and Bossaerts (2009) and Murawski and Bossaerts (2016).\textsuperscript{14} Importantly, the presence of markets does not invalidate the correlation between Sahni-$k$ and computational performance.

We used the number of items participants moved into and out of the knapsack as a proxy for effort. The mean number of moves was $XX$ (min = $x$, max = $x$, $SD = x$). Correct and incorrect items were moved into and out of the knapsack equally frequently ($t$-test, $p < x$). To test whether the number of moves depended on instance complexity, we related the number of moves a participant made, to Sahni-$k$ of the instance (mixed-effects model with a fixed effect for Sahni-$k$ and random effects for participant and session on the intercept). We found no effect of Sahni-$k$ ($p > x$).

Participants traded 12 shares in markets corresponding to correct items (items in the solution) and 13 shares in markets for incorrect items on average. There was little difference in trading across sessions and instances. More difficult instances (as measured with Sahni-$k$) did not generate more trade (see SOM for details).

In the remainder of this section, we report several tests of EMH. We first

\textsuperscript{13}Note that the number of sets the algorithm considers increases exponentially in $k$. For an instance with 12 items, the number of sets considered if $k$ is equal to 1, is 1 whereas if $k$ is equal to 4, the number of sets considered equals 20,736.

\textsuperscript{14}For a comparison of Sahni-$k$ as a measure of instance complexity with other measures, see Murawski and Bossaerts (2016).
consider the question whether security prices revealed security values, that is, whether they revealed the optimal solution of an instance. This is a direct test of EMH. We then examine whether participants who found the optimal solution were able to profit from their information. This is an indirect test of EMH. Finally, we examine whether uninformed traders benefitted from the information revealed by security prices.

**Issue 1: Did Securities Prices Reveal The Optimal Solution?**

To evaluate whether prices correctly reflected the solution of an instance, we constructed a “market performance” metric as the distance, in item space, of “market solutions” implied by trade prices from the optimal solution. Distance in item space is measured as a score which is incremented by one point if a correct item was in the proposed knapsack, and decremented by one point if an incorrect item was left out, and zero otherwise. The score is subsequently scaled by dividing by the score of the correct KS. To construct “market solutions,” we interpreted average trade prices as the market’s “belief” that the item belonged in the optimal solution. We then bootstrapped a “market solution” by drawing items based on these beliefs and filling the knapsack until capacity was reached. We repeated this process 10,000 times per instance, each time computing the performance score, and subsequently averaging across repetitions. If prices perfectly revealed an instance’s solution, we would draw only from correct items, and hence, obtain a perfect performance metric.\(^{15}\)

Information revealed in average market prices allowed us to achieve recover the optimal solution in 96.4% of bootstrapped attempts in case of the instance with Sahni-\(k\) equal to 0. The proportion decreased monotonically to 85.6%, 78.3%, 75.8% and 72.5% as Sahni-\(k\) from 1 to 4. We conclude that market prices were never efficient in the sense of EMH as they never revealed the optimal solution with perfect accuracy.

\(^{15}\)In the SOM, we present results from an alternative method where market performance is computed directly from average trade prices. This method confirmed our findings.
We compared the performance of the “market knapsack” to the performance of the knapsacks submitted by individual participants. We constructed performance scores for individuals the way we did for the market. We found that in four of the five instances, the market knapsack performed better than the knapsack submitted by the average participant (Fig. 2). Performance of the market decreased with difficulty at the same rate as for individual participants.

**Figure 2. Performance Of Instance Solutions Generated From Average Trade Prices Compared To Individual Submissions.**

*Note:* Market performance score, measured as distance in item space of “market knapsack” obtained from simulations based on average trade prices (red), against average participant score, measured as distance in item space of submissions (blue), across all sessions, stratified by instance difficulty. Error bars denote standard errors across sessions. Dashed line is regression line (Pearson correlation $r = 71.4\%$).

Note that except for the most difficult instance ($Sahni-k = 4$), there was always at least one participant who found the optimal solution (in the instance where $Sahni-k$ was equal to 0, the majority of participants found the solution). This means that the information to value securities correctly was available among
participants. Yet market prices never allowed us to construct a “market knapsack” with a perfect performance. We conclude that EMH did not hold. The market performed better than individual participants, but prices did not reflect the value of securities.

**Issue 2: Do Informed Traders Make Money?**

To perform our second test of EMH, we correlated individual performance in an instance with the respective earnings (in Australian dollars) from trading in the marketplace. We compute the former as distance in item space of submitted knapsack from the correct solution, as described above.

The average payoff in an instance for informed participants (those who found the solution of the instance) was $8.22, compared to $5.11 of uninformed participants. Moreover, in every instance, the highest payoff among all participants was earned by an informed participant.\(^{16}\)

Graphical presentation and a formal test confirm the economic magnitude and statistical significance of the effect of superior computational performance (Fig. 3). A ten percentage point increase in the computational performance score was associated with additional earnings of 70 cents. We formally verified the effect with a generalised linear mixed effects model. The main regressor was the performance score, with subject-level fixed effects, and controlling for instance difficulty (Sahni-\(k\)) and experimental session.

Considering earnings and performance in individual experimental sessions, we found that they were highly correlated: an increase in performance score of 10 percentage points increased earnings in the marketplace by almost $4 on average (Fig. 4). Note that only two participants received a perfect score, implying that only two participants solved all five KS instances correctly.

\(^{16}\)The average maximum payoff was $21.05 in the instance with Sahni-\(k\) equal to 0, compared to $12.50, $13.70, $16.25, and $11.95 for instances with Sahni-\(k\) equal to 1 to 4, respectively. Notice the non-monotonicity of earnings. This is related to the way we allocated shares, which as mentioned before, attempted to balance incentives to trade and fairness. Average earnings were highest in the instance with Sahni-\(k\) equal to 0 and lowest in the instance with Sahni-\(k\) = 2. Average earnings amounted to $8.41, $5.13, $5.03, $5.93 and $7.08 for Sahni-\(k\) 0 to 4, respectively.
Figure 3. Individual Trading Payoff Per Instance Against Distance From Solution Of Submitted Knapsack.

Note: Individual payoff from trading for each instance against performance score for the instance; performance score is measured as distance in item space of submitted knapsack. Slope coefficient (fixed-effects regression; see text) equals 7.01 ($p < 0.001$); Pearson correlation $r = 0.303$. 

Figure 4. Individual Trading Payoff Per Session Against Average Distance From Solution Of Submitted Knapsacks.

Note: Individual session earnings from trading against average performance score for the session; performance score is measured as distance in item space of submitted knapsacks for the five KS instances. Slope coefficient (fixed-effects regression; see text) equals 37.1 ($p < 0.001$); Pearson correlation $r = 42.2\%$. 
Therefore, our comparison of the earnings of informed against uninformed, and the analysis of the effect of superior computational performance both confirm that the EMH did not hold: better-informed traders made significantly more profit; and participants whose submitted knapsacks were closer to the optimal solution (in item space) generated higher earnings from trading.

**Issue 3: Do Uninformed Traders React To Information Reflected In Prices?**

In the rational expectations equilibrium used to provide empirical content to EMH, the uninformed are assumed to know the mapping from states to prices and use this mapping to infer states from observed prices. In the traditional model, a “state” refers to the average of the bits of information in the economy; here, it should be interpreted as the correct solution to the KS instance at hand.

We investigated whether uninformed indeed “read” information from prices. Specifically, we study to what extent trade induced uninformed (those that did not submit the correct solution) to re-visit and improve their knapsack, moving it closer to optimum in item space.

Prior research has shown that an important reason why humans may not find the optimum is because they tend not to re-consider incorrect items that they put into the knapsack early on (Murawski and Bossaerts, 2016). Poor episodic memory may explain this reluctance to re-visit early moves. Here, we ask whether trade in such items made it more likely that uninformed traders took them out.

Specifically, we ran a generalised linear mixed-effects logit model on the probability of removing an early inclusion of an incorrect item, and tested whether trade price and volume and their interaction had the effect of increasing this probability.\(^{17}\) We decided that an item was included “early on” when the participant moved it into the knapsack within the first two minutes of trading.

Results confirmed that the probability of removing an incorrect item placed into the knapsack early increased with low prices (coefficient: -3.491, \(p < 0.001\)).

\(^{17}\)Model selection analysis based on the Bayesian Information Criterion suggested that a model without fixed effects fit better.
The effect of a 10% change in prices is an 11% increase in the removal probability. The effect is stronger when the security corresponding to the item is traded more heavily (interaction term coefficient: 13.725, $p = 0.027$). However, there was no significant effect of trading volume itself ($p = 0.068$).

Fig. 5 provides a graphical display of the effect of prices on the probability of removal of an item that was incorrectly put in the knapsack during the first two minutes of trade.

![Graph](image)

**Figure 5. Impact Of Trade Prices On Removal Of Incorrect Items**

*Note: Fraction of incorrect items moved into knapsack within two minutes of trading that are eventually removed, as a function of average trade price.*

These findings suggest that trade, combined with low prices, induce participants into re-considering faulty moves, thus improving overall computational performance, and hence, securities valuation.

We did not find an effect on KS choice from trade in high-priced securities.
We conjecture that short-sale restrictions contributed to the asymmetry between high-priced and low-priced securities: if an item is deemed to be overpriced, participants can only sell shares that they already own, and hence cannot put more pressure on prices.

III. Discussion

We report results from a markets experiment aimed at testing EMH when heterogeneous information emerges spontaneously due to computational complexity. We found that prices did not reveal security values, and that informational efficiency deteriorated as computational complexity increased. Nevertheless, prices could be used to construct candidate solutions (knapsacks) that were closer to the optimal solution than those that participants submitted on average. We also document that informed traders (those who submitted the optimal solution) earned significantly more from trading on their information, and that uninformed traders (those who did not submit the correct solution) earn more the closer their submission was to the solution. Closeness was measured in terms of the number of correct items in the submission and the number of incorrect items left out of the submission. Finally, we discovered that prices, in conjunction with trade, fed back into problem solving: low prices for heavily traded securities induced uninformed traders to re-consider incorrect items they had put in their knapsack early on. Consequently, while EMH did not hold, a core principle of the traditional theory behind EMH was upheld, namely, that uninformed “read” information from prices.

Tests of EMH require one to correctly adjust for risk (Fama, 1991). Controlled experiments, like ours and pioneering studies in this area (Plott and Sunder, 1982, 1988), do not suffer from this shortcoming. We did not have to extract “risk adjusted” returns because all participants were equal ex ante, in terms of allocations, knowledge and attitudes towards risk, uncertainty and complexity. (We hasten to add that notions like risk and ambiguity aversion are ill-defined
in our setting since there is neither risk nor uncertainty, and the true meaning of complexity aversion, to our knowledge, is yet to be explored.)

To conceptualise complexity, we used standard notions from computer science. These notions were developed based on a theoretical computational model (Turing machine), and it was not \textit{a priori} obvious that they would extend to human computing. Recent evidence suggests that complexity theory does extend to human decision-making (Murawski and Bossaerts, 2016), and the findings reported here corroborate this notion. Therefore, evidence is mounting that the complexity theory is universal, a conjecture known as the Church-Turing thesis (Church, 1934; Turing, 1936).

Here we show that complexity theory extends even to markets: we found that prices violated EMH more when valuation had higher computational complexity, where complexity was measured in terms of Sahni-\textit{k}. This metric provides a measure that aptly summarises human performance in solving the KS problem (Murawski and Bossaerts, 2016). Our finding confirms earlier experiments on the use of markets as a mechanism to solve computationally complex problems (Meloso, Copic and Bossaerts, 2009). As such, markets can be thought of as “computers” as well.

We divided our participants into a group of “informed” and “uninformed” traders, based on whether they submitted the optimal solution. It is important to realise that informed traders may actually not have been aware that they knew the solution. The nature of the KP is such that the only way to ascertain that a candidate solution is the optimal solution, is to solve the instance. Evidence from prior studies suggests that even when participants submit the optimal solution, they are often not aware that their solution is indeed optimal (Meloso, Copic and Bossaerts, 2009; Murawski and Bossaerts, 2016).\textsuperscript{18}

Given that our experiment only had 18-20 participants per trading session, it

\textsuperscript{18}It is highly unlikely that humans are able to implement the algorithms to compute the optimal solution, such as dynamic programming or brute-force search (Murawski and Bossaerts, 2016).
could be argued that our results are due to insufficient competition or liquidity. However, other experiments have shown that EMH can be obtained in settings with fewer participants. However, in those experiments, valuation only required simple aggregation of heterogeneous information (through averaging). We show that EMH does not hold when security valuation is computationally complex.

In the context of market efficiency, the Grossman-Stiglitz (GS) Paradox (Grossman and Stiglitz, 1980) is the biggest obstacle to information production. The GS Paradox refers to the inability of agents to recuperate the cost of information gathering when market prices satisfy EMH. Our experiments clearly show that better-informed traders earned more, and hence, could have recuperated costs if we had charged them. As such, the GS Paradox does not extend to a situation where information heterogeneity stems from computational complexity.

Importantly, we found that uninformed traders learned from prices. This is in sharp contrast with Asparouhova et al. (2015). There, correct security valuation required one to solve a computationally simple but highly non-intuitive Bayesian problem (variations on the so-called Monty Hall problem). When faced with prices that were at odds with their beliefs, participants did not learn from prices, but instead retreated from exposure to risk by trading to portfolios with valuations that did not depend on the Bayesian problem at hand. Here, we reported how uninformed traders got cues from more heavily traded, low-priced securities, nudging them to re-consider whether to delete the corresponding items from their knapsack.

Feedback from trading to problem solving may explain why earlier experiments showed that markets can be used to help people solve KS instances. Meloso, Copic and Bossaerts (2009) report that more participants managed to find the correct KP solution when incentives were similar to those in our experiment, namely, one had to trade to make money. The benchmark incentive scheme in that study was one were only the first to submit the optimal solution would be paid a fixed prize. This prize was significant: it amounted to the sum total of dividends paid in the
markets treatment. Fewer participants found the correct solution under the prize treatment.

We discovered that markets mitigate one of the strongest biases that keep individuals from discovering the correct KP solution, namely, hesitation to re-consider items that were added to the knapsack early on (Murawski and Bossaerts, 2016). In the experiment reported here, pricing, in conjunction with trade, made participants re-visit parts of the solution that they had constructed within the first two minutes of trading. It is interesting to note that this explanation differs from the conjecture in Meloso, Copic and Bossaerts (2009) as to why markets cause better problem-solving. There, it was suggested that success in a prize system hinges on the belief that one is good enough to sometimes be the best, a belief that most people would never entertain. In contrast, in markets, one merely has to believe to be better than the median, something the majority does believe, a situation known as the “overconfidence bias” (Kahneman and Tversky, 1977). As such, participants worked harder in the markets treatment. Here, we showed that there was an effect beyond mere market participation, namely, trading activity provides valuable cues that improved individual problem solving.

Altogether, our findings demonstrate that markets may provide more powerful incentives to solve complex problems than a prize system. The prize system in Meloso, Copic and Bossaerts (2009) is analogous to the current patent system, and hence, our findings are relevant for the current debate on the desirability of patents as a way to promote innovation. Indeed, intellectual discovery can be thought of as the solution of a combinatorial problem such as the KP (Boldrin and Levine, 2002). A recent empirical study corroborates this claim, by showing that patents filed with the U.S. Patent Office between 1790 and 2010 were mostly for inventions that combined existing technologies in novel ways rather than opening up fundamentally new avenues of exploration (Youn et al., 2015). Our findings suggest that markets may promote innovation better than a prize system.

To illustrate this notion, consider markets in Li, Na, Mg and other chemicals
(we are using standard chemical abbreviations). They are potential components of future battery technology. To determine which chemicals markets to invest in requires one to assess which component, or maybe combination of components, will be required for the best battery technology. This casts intellectual discovery squarely in terms of the KP and the markets we designed to profit from finding the optimum. Inventors are induced to participate in the marketplace, and earn money by buying those components that they believe are in the best battery technology, while selling others. If EMH held, inventors would not be able to make money from their knowledge. Other incentives, like prizes or patents, would be needed to incentivise discovery. However, we found that EMH does not hold in such a situation. Hence, markets can provide incentives to innovate.

This conclusion is important for economic history as well. It is generally accepted that the patent system provided the main impetus for technological advances over the last century and a half. However, at the same time markets penetrated all parts of life, and it may very well have been that markets were the major facilitator of innovation rather not patents. This conjecture agrees with historical evidence that technological advances can be far bigger during epochs with share trading but without patents than in epochs with only patents (Nuvolari, 2004).

The failure of prices to reveal all available information resolves another problem with EMH, namely, the “Hirshleifer Effect” (Hirshleifer, 1971). This is the detrimental effect on welfare that agents experience when they happen to be endowed with assets that prices reveal to be of little value. In the battery example above, if Mg emerges as useless for the best technology and this is immediately revealed in prices, agents endowed with Mg have poor terms of trade from the beginning. They would have wished to be able to take out insurance before initial endowments were revealed.

The Hirshleifer Effect is to be counted with in traditional experiments on EMH, where bits of information are spread across participants, but when averaged, pro-
vide the best estimate of the value of the securities at hand. Participants end up with vastly different total payoffs (and will complain) unless the experimenter deliberately rigs the initial endowments to generate equal *ex ante* earnings. For more discussion, see, e.g., Asparouhova and Bossaerts (2016). Here too, we assigned initial allocations in “fair” ways, to guard against participant dissatisfaction with our experiment in case EMH were to obtain and the Hirshleifer Effect were to bite. Because EMH did not obtain, participants did not complain, and with hindsight, we could have dispensed with the complication of ensuring “fair” initial endowments.

Our findings allow one to put into perspective the evidence on EMH from historical analyses of field data. Fama (1991, 1998) surveys a vast body of studies that appears to suggest that EMH holds, because anomalies could be attributed to chance (sometimes the null will be rejected) or to methodological errors. Behavioural finance scholars reject this conclusion, starting with Bondt and Thaler (1985), who claimed that long-run price reversals are caused by overreaction. Among others, Hirshleifer (2001) summarises the evidence against EMH.

Our experiment should shed new light on the EMH controversy. When heterogeneous information emerges because agents hold dispersed information that, when averaged, provides strictly better information, markets can be expected to satisfy EMH. That is, EMH holds under “finite sample complexity.” However, when computational complexity is the cause of heterogeneous information, EMH will not obtain. In the field, it is not obvious whether the situation is one of finite sample complexity or of computational complexity; maybe even both. Future research could juxtapose two historical field cases where one can unambiguously be identified as finite sample complexity while the other is a case of computational complexity. To the extent that our study holds external validity (and those on EMH under finite sample complexity, such as Plott and Sunder (1982, 1988)), we expect to first case to produce evidence in favour of EMH, while EMH would be rejected in the latter case.
The evidence in favour of EMH from earlier laboratory studies has led to the emergence of a new type of information aggregation device in the field, namely, prediction markets. These markets are purposely designed to aggregate the information “that is out there,” i.e., to harvest the “wisdom of the crowds.” Successful field implementations abound (Arrow et al., 2008), but instances have emerged where market prices predicted the wrong outcome, such as for the Brexit vote in the U.K. in 2016. Fundamentally, the power of prediction markets depends on the validity of EMH. Our research suggests that prediction markets will serve their purpose if the situation is one of finite sample complexity, but not under computational complexity.

American Economic Review Pointers:

- Do not use an "Introduction" heading. Begin your introductory material before the first section heading.
- Avoid style markup (except sparingly for emphasis).
- Avoid using explicit vertical or horizontal space.
- Captions are short and go below figures but above tables.
- The tablenotes or figurenotes environments may be used below tables or figures, respectively, as demonstrated below.
- If you have difficulties with the mathtime package, adjust the package options appropriately for your platform. If you can’t get it to work, just remove the package or see our technical support document online (please refer to the author instructions).
- If you are using an appendix, it goes last, after the bibliography. Use regular section headings to make the appendix headings.
- If you are not using an appendix, you may delete the appendix command and sample appendix section heading.
Either the natbib package or the harvard package may be used with bibtex. To include one of these packages, uncomment the appropriate usepackage command above. Note: you can’t use both packages at once or compile-time errors will result.

IV. First Section in Body

Sample figure:

Figure here.

Figure 6. Caption for figure below.

*Note:* Figure notes without optional leadin.

*Source:* Figure notes with optional leadin (Source, in this case).

Sample table:

<table>
<thead>
<tr>
<th>Heading 1</th>
<th>Heading 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row 1</td>
<td>1</td>
</tr>
<tr>
<td>Row 2</td>
<td>3</td>
</tr>
</tbody>
</table>

*Note:* Table notes environment without optional leadin.

*Source:* Table notes environment with optional leadin (Source, in this case).

References here (manual or bibTeX). If you are using bibTeX, add your bib file name in place of BibFile in the bibliography command.

REFERENCES


Arrow, Kenneth J, Robert Forsythe, Michael Gorham, Robert Hahn, Robin Hanson, John O Ledyard, Saul Levmore, Robert Litan, Paul


Mathematical Appendix