

The Oracle or the Crowd? Experts versus the Stock Market in Forecasting Ceasefire Success in the Levant

Gerald Schneider/Maya Hadar/Naomi Bosler
Department of Politics and Management
University of Konstanz

The forecasting literature has come to mistrust the predictions made by experts who forecast political events in mass media. Distinguishing between judgments made by one or few individuals ('oracles') and assessments made by larger groups ('crowds'), we contrast journalistic predictions with forecasts stemming from the financial industry. These two competing views were evaluated in a quantitative analysis of the *ex ante* success of 24 ceasefire agreements in various conflicts which took place in the Levant from 1993 to 2014. Our analysis compares the forecasts appearing in press commentaries (*Haaretz*, *Jerusalem Post* and *New York Times*) with the expectations that the Tel Aviv Stock Exchange had about the stability of these cooperative efforts. To evaluate the predictions of these very dissimilar sources, the effectiveness of the ceasefires were analysed through the number of violent events following the official start of the truce. The analysis shows that the financial industry performs better than the media industry in the comparative evaluation of ceasefire forecasts, but that neither source provides sufficiently accurate predictions. The partial support for the crowd thesis is discussed in light of a recent literature that resuscitates the usage of well-trained experts for forecasting purposes, but warns against the dramatizing predictions of media pundits.

Word count: 8830

Forthcoming: *Journal of Peace Research* 54(1), 2017

Corresponding author: Gerald.Schneider@uni-konstanz.de

Introduction

Forecasting is a booming business (Sherden, 1997; Bueno de Mesquita, 2009). It remains, however, both practically and academically unclear who should fulfil the ever-lasting wish of policy makers and the public to know the future before it has happened. This article contrasts predictions of political violence produced by two important economic sectors– the media and the financial industry. While the former provides its audience with forecasts about political developments through commentaries and editorials, the latter generates predictions through the valuation of politically sensitive assets.

Evidently, the judgements made by journalists and stockbrokers on particular political developments feed into each other. However, the appraisals of the two industries differ in important ways. First, mass media generate direct forecasts through their more or less explicit predictions of how a certain process will evolve. The financial industry, by contrast, only generates indirect predictions through the assessments of how political events affect the returns on particular investments. Second, while the accuracy of a forecast directly impinges on the income and possibly the job security of financial traders, failed predictions will only exceptionally bring havoc to the career of a media commentator. Third, whereas the financial industry acts as an anonymous and often fairly large group, editorial and commentaries are typically written by a specific newsperson or by a small team of journalists.

For the past decades, the social scientific literature has debated the question of whether individual experts ('oracles') or groups of laypeople ('crowd') provide better forecasts. As two non-fiction bestsellers show, both positions enjoy widespread popularity. Taleb (2007) warned at the outbreak of the Great Recession in a polemic that the financial community had collectively downplayed the risk of financial crises and thus had not taken sufficient precautions against what he calls 'extremistan'. His rant about the industry's mistaken adherence to the normal distribution for forecasting crises echoes Le Bon's (1895) infamous thesis that the masses are foolish and only a few enlightened individuals, if anyone at all, are able to correctly augur a looming catastrophe.

The position that 'oracles' predict the future more precisely than groups of highly paid forecasters in the finance industry is in contrast to the 'crowd' conjecture, popularized by Surowiecki (2004) and advocated at least since Condorcet's (de Caritat, 1785) establishment of the Jury Theorem and Galton's (1907) classic work on

belief aggregation.¹ According to this alternative viewpoint, large groups of laypeople do much better in ascertaining the truth than individuals who possess ample personal knowledge in a particular area. Tetlock's (2005) path-breaking study on the accuracy of expert forecasts, further supported the scepticism towards the usefulness of expert knowledge, although his most recent work and related studies qualify this position (Mandel & Barnes, 2014; Mellers et al. 2014; Tetlock, 2005; Tetlock & Gardner, 2015).

We examine these contradictory expectations through an evaluation of the accuracy of newspapers commentaries and financial markets in predicting the *ex ante* success of ceasefires in the Levant. Sadly, the Middle East is a region which has been plagued by conflict over the past 60 years. Israel has rarely, if at all, experienced a period of peace in its short history. Over 20 ceasefires have been declared since the end of the Cold War alone, either unilaterally or jointly, and all have, ultimately, failed to bring permanent peace to the region. The 24 ceasefires examined here, concluded at various times between Israel and Hezbollah, Hamas, the Palestinian Liberation Organisation (PLO) and several other Palestinian groups between 1993 and 2014, have achieved varying degrees of relative calm.

It is against this background that the current study is conducted. Guided by Fleischer & Buccola (2002), we propose that the Israeli tourism industry listed on the Tel Aviv Stock Exchange (TASE) should be especially sensitive to the success or failure of ceasefire agreements. This theory is supported by Neumayer (2004) who finds strong evidence that human rights violations, conflict and political violence have a negative effect on tourist arrivals. Furthermore, we expect that, whilst export and import sectors suffer from war and conflict, certain industries, such as the defence sector, may in fact profit from such events (e.g. Schneider & Troeger, 2006; Guidolin & La Ferrara, 2007). We extend these expectations regarding the redistributive effects of violent events to hypothesize that financial traders will carefully evaluate whether a ceasefire will hold, as the success or failure of such an agreement will directly affect the returns on specific investments.

¹ The Condorcet Jury Theorem, in its canonical version, states that the competence of a group in recognizing the true state of the world is larger than the average individual competence and that this collective capability approximates 1 if the size of the committee grows (for a generalization, see List & Goodin, 2001). Galton (1907) empirically demonstrated the advantages of collective belief aggregation through his systematic analysis of a 1906 livestock fair tournament, where the mean and median guesses of the weight of an ox were more accurate than any individual estimate.

The predictive power of the stock market will be compared to that of newspaper commentators in forecasting the success or failure of ceasefire agreements. We acknowledge that journalists, with their need to dramatize their stories and to please an often biased audience, rather resemble what Tetlock (2005), drawing on Berlin (1957), has called ‘hedgehogs’ – forecasters whose predictive capacities are underdeveloped despite their expertise in the domain for which the predictions were produced. Yet, as both policy makers and the media audience will continue to pay attention to pundits, we believe that juxtaposing the forecasts of the media and the financial industry is an endeavour of considerable theoretical and practical importance. This is particularly the case as other competitive evaluations of forecasts compare different types of experts or intelligence reports (Mandel & Barnes, 2014; Mellers et al. 2014; Tetlock, 2005; Tetlock & Gardner, 2015), rival game-theoretic models (Thomson et al. 2006; Schneider, Finke & Bailer, 2010; Bueno de Mesquita 2011) or varying econometric specifications of the same time series or panel of cases (Brandt, Freeman & Schrodt, 2011, 2014; Rost, Schneider & Kleibl, 2009), but hardly ever contrast predictions obtained through different sources. That is despite Meehl’s (1954) pioneering study comparing clinical and statistical predictions.

To evaluate substantively different *ex ante* predictions of the same processes, we have conducted content analyses of three quality newspapers (*Haaretz*, *Jerusalem Post*, *New York Times*) and event studies of two assets (*Dan Hotels*, *Imco*). The analysis suggests that we can use carefully selected stock market indices to predict political developments that are economically important. While a financial asset from the tourism sector, *Dan Hotels*, is a better predictor of ceasefire success than media experts at the *New York Times*, *Haaretz*, and the *Jerusalem Post*, an asset from the Israeli defence sector, *Imco*, largely fails in forecasting the fate of ceasefires. In other words, the evidence thus rather supports the ‘crowd’ over the ‘oracle’-thesis. However, neither source provides truly satisfactory predictions.

Theory and literature

There is a growing literature that advocates the usage of advanced predictive techniques to improve the empirical evaluation of theoretical claims and of the early warning schemes used in foreign policy decision making (Schneider, Carey & Gleditsch, 2011). Yet, the predictive accuracy of a particular methodology, as sophisticated as it might be, further depends on the quality of the information that

feeds the forecasts. This caveat is particularly important for judgmental data that are based on the expertise of individuals or groups.

Tetlock (2005, 2010) shows that expert political judgement is not necessarily a reliable method of producing accurate predictions. Forecasters fall easily into the trap of overconfidence whenever they are not accountable for their judgements, if strong prior beliefs guide their assessments and if the time horizon for the predictions extends over several years.² Newer experimental research in Social Psychology shows that the identification of good forecasters, the sharing of information as well as systematic training might resuscitate the role of individual experts (Mellers et al. 2014). Tetlock & Gardner (2015: 74) argue that the main advantages of individuals with a high accuracy in predicting events – ‘foxes’ or ‘superforecasters’ to use their terminology – is their ability to aggregate information across different sources, making them comparable to Galton’s crowd: ‘They integrate perspectives and the information contained within them. The only difference is that the process occurs within one skull’. Mandel & Barnes (2014: 10987) conclude in an evaluation of strategic intelligence reports that ‘tempered optimism’ is warranted. Accordingly, well-calibrated expert predictions ‘explained 76% of the variance in geopolitical outcomes.’ As we do not have access to such files, we examine the forecasts of media commentators on the anticipated success of truces in the Levant.

The new socio-psychological evidence suggests that journalists or experts, interviewed by them, will almost by necessity not belong to the super-forecasters. As Tetlock & Gardner (2015: 72) maintain, ‘editors, producers, and the public ... go looking for hedgehogs, who just happen to be bad forecasters’. The pressure to come up with a dramatic assessment or an opinion that lives up to the prejudices of the audience of a particular media outlet risks transforming these commentators into oracles. Note that our evaluation is confined to the subgroup of experts who act in the public and, for better or for worse, shape public opinion with regards to the success or failure of ceasefire agreements as well as other important political developments. This explicit focus on journalistic opinions is the main difference between this article and others which analyze news through word counts and other techniques to predict the risk of armed conflict (e.g. Chadeaux, 2014).

² In contrast to these assessments Bueno de Mesquita (2009, 2010) maintains that this skepticism is not warranted for his game-theoretic forecasting approach. The main advantage of such predictions is that the level of expertise required from an interview partner is modest and only relates to an evaluation of the present.

H1: The more optimistic the reaction of media experts to the announcement of a ceasefire is, the more successful the ceasefire will be.

The main reason as to why financial market information should improve forecasts for some political developments depends on the costliness of the decisions that traders make. A broker incorporates relevant economic *and* political information when deciding upon selling or buying a particular security. If a financial market heavily depends on political decisions, traders build up expectations about the likelihood of certain scenarios in the political realm. Stock markets have systematically responded to escalatory and de-escalatory moves in three conflicts throughout the 1990s (Schneider & Troeger, 2006; Brune et al. 2015) and to the deliberations of the European Council on security policy making (Bechtel & Schneider, 2010). As some economically important events, such as upcoming meetings between heads of states, are easily foreseeable, it would be irrational for traders to refrain from partly anchoring their current buying and selling decisions on them. To put it differently, if traders refrain from building sensible forecasts with regard to important political developments, they not only jeopardize the value of the securities they are trading with, but also possibly harm their personal careers. Obviously, markets can be wrong in their assessments, and they might have, if Paul Samuelson's famous dictum 'Wall Street indices predicted nine out of the last five recessions' (*Newsweek*, Science and Stocks, 19 Sep. 1966) holds, a tendency to overpredict.

Asset prices reflect the collective wisdom that the market holds on how certain political events affect the returns on a particular investment. Media commentaries, by contrast, are individual assessments, although the commentators are undoubtedly also concerned with the extent to which a certain opinion deviates from the beliefs and prejudices of the media audience. In the following, we assess whether stock market data can be used to forecast the success of ceasefires in the Levant. This conflict, which has been dominated by the confrontation between Israel and the Palestinians throughout the past six decades, certainly belongs to the category of disputes that have general economic consequences (Schneider & Troeger, 2006) and also affect individual sectors (e.g. Fleischer & Buccola, 2002). As developments within this conflict region are well reported, several researchers have already attempted to forecast escalations for the various dyads (e.g. Brandt, Freeman & Schrodtt, 2011, 2014). This article moves beyond these event data-based time series models in order

to produce predictions by including information from a quasi-expert collective – the stock market.

To the authors' knowledge, only a limited number of studies has tried to use financial market information to predict violent conflicts. Bueno de Mesquita (1990), descriptively uses the money market discount rate to evaluate the expectations contemporary agents had over the Seven Weeks' War between Austria and Prussia and the crucial battle at Königgrätz. Moreover, Schneider (2014) uses aggregated stock exchange information to predict *ex post* the ups and downs of the conflict between Israel and the Palestinians.

The theoretical background which enables us to consider the stock markets as collective *ex ante* forecasters is the efficient market hypothesis (EMH) according to which 'a market in which prices always 'fully reflect' available information is called "efficient"' (Fama, 1970: 383). In other words, the EMH claims that the price of an asset reflects all relevant information at a given point in time. It is important to note that not all market participants need to be well informed for the EMH to work. Indeed, it is sufficient that only a few traders are well informed, assuming that the errors made by non-informed traders cancel each other out.³

One can also expect the stock market to react to the announcement or the anticipation of a ceasefire, and to develop its own expectations as to the probability of its success. However, the effect of a successful ceasefire agreement is not homogenous across sectors. Firstly, empirical evidence supported by sound theory would suggest that the tourism industry is negatively affected by conflict. Fleischer & Buccola (2002) find that foreign tourist demand for Israeli hotel accommodation is moderately sensitive to terrorist attacks and that local tourism provides little buffer for reductions in foreign tourist arrivals. In a cross-country study of tourist arrivals between 1977 and 2000, Neumayer (2004) demonstrates that human rights violations and conflict had a negative effect on the tourist industry.

³ One application of the efficient market hypothesis to the area of political forecasting is found in prediction markets, in which the payoff one receives on a stock is tied to the outcome of a future event. By binding the return on the asset to an event that may or may not occur in the future, the price of the asset can be interpreted as the market's expectation of the likelihood that the event will occur (Arrow et al. 2008). However, they only allow the forecasting of events for which contracts are directly available. Furthermore, the development of a reliable prediction market is time consuming and relies on the assumption that the market garners enough interest to attract a large body of buyers.

H2a: The stronger the rally of tourism stocks in anticipation of the announcement of a ceasefire is, the more successful the ceasefire will be.

Secondly, the literature is also united in its belief that the defence and security sector should benefit from increased tensions, therefore suffering under peace. Berrebi & Klor (2005) establish that terrorist events in Israel have a significant and positive effect on Israeli defence as well as on security stocks traded in American markets. In addition, Camyar & Ulupinar (2015: 641) demonstrate that war ‘exercises a positive, but surprisingly weak, distributive effect on the U.S. arms sector.’

H2b: The stronger the fall in defence stock prices in anticipation of the announcement of a ceasefire is, the more successful the ceasefire will be.

Finally, if one takes the EMH a step further and notes that the stock market, being information efficient, should take into account *all relevant information*, that is, including the information conveyed by the media, then it is logical to expect that the stock market should provide a better forecast of the likelihood of a truce’ success than experts do. This logic is also supported by the empirical evidence of Nofer & Hinz (2014) in the analysis of European stock prediction communities versus expert assessments.

H3: The stock market is a better predictor of ceasefire success than are media experts.

Methodology

The Levant offers a unique opportunity to study the effect of conflict on the stock market. The Tel Aviv Stock Exchange (TASE), founded in 1953, is a mature and highly capitalized market that is likely to be sensitive to changes in the Arab-Israeli Conflict. Therefore, we will consider the body of ceasefires occurring between Israel and any other party in the period 1993-2014. The term ‘ceasefire’ is defined here as the explicit agreement to cease violent attacks against the other party. There are certainly cases where the agreement is implicit or *de facto*; however, these are excluded from this study as they lack a date of announcement leading to intangible stock-market and press reactions.

The list of ceasefires was built using the Penn State Event Data Project (PSEDP)'s CAMEO data set for the Levant. After removing all events in which Israel was not involved or that occurred before 1993, the events coded as ceasefires were selected. This results in a body of 324 events; which we then went through one by one to verify whether it was an actual ceasefire or incorrectly coded⁴. The fact that the actual body of ceasefires was reduced to 24 during this process justifies our high-level of caution when using machine-coded data.⁵

In the case of the Levant, the operationalization of the dependent variable, ceasefire success, is even more problematic than it might be elsewhere. This is due to the special nature of the Arab-Israeli conflict. Ceasefires that have occurred in the past two decades between Israel and a second party tend to be less durable and far less strictly observed than in other regions. For example, whilst most conceptualize a ceasefire as a total halt to all violent events between the two parties, most of the 24 ceasefires covered a period in which violence still continued at least sporadically. Due to this phenomenon it becomes increasingly difficult to conclude when the truce ended and when full scale hostilities were renewed. For this reason, we employed the number of violent events occurring within 7 and 14 days, respectively, from the official start of the truce in order to measure the success of a ceasefire.⁶

The data for the dependent variable was gathered from the same source used to define the body of cases, i.e. the CAMEO data set for the Levant region of the Penn State Event Data Project (PSEDP).⁷ As we are working with overdispersed count data, a negative binomial model will be employed.

The predictive power of experts: The first section of our analysis is designed to estimate the ability of experts, specifically journalists, to predict the success of

⁴ The CAMEO data was used as a starting point to identify ceasefire-related articles, which were then individually examined to determine whether they could be classified under the above-noted definition of ceasefire. In addition, we cross checked the data using secondary sources such as the BBC timeline for the Middle East and The Israeli Defense Force spokesperson unit's website. The latter were used to identify the most current ceasefire events (event numbers 22-24), which were out of the time range covered by CAMEO. The original body of 324 commentaries relating to the ceasefire declarations (1993-2014) included several duplicates.

⁵ The events included in the AFP Cameo data base were coded by the software TABARI,(Text Analysis By Augmented Replacement Instructions).

⁶ A time period of seven days was also applied. See the Online appendix for further details.

⁷ From the original data set of 223,608 events, we first filtered out all events in which Israel or an Israeli agency was not a participant. We then deleted all events except those which were violent in nature. Finally, we counted the number of events occurring in the period after the ceasefire date and saved these counts as the dependent variable.

ceasefire agreements. For the purpose of the analysis, three newspapers, the Hebrew version of the liberal *Haaretz*, the conservative *Jerusalem Post* and the liberal *New York Times*, have been chosen for their reputations as reliable dailies with a broad readership and with different ideological positions on the conflict between Israel and the Palestinians.⁸ Issues of the newspapers released on the day that the truce was announced and the following day were searched for articles referring to the ceasefire.⁹ The relevant articles have been downloaded from *Nexis* (*JP*, *NYT*) or retrieved from Israeli archives in the form of microfiches (*Haaretz*).

The coding instructions have been developed so that the coder can execute the content analysis, which constitutes nine sequential steps, without any further knowledge of the topic. Each individual step articulates a clear task that the coder is required to perform and offers some guidance as to how to do so. We have taken steps to ensure that the process is as replicable as possible. Firstly, the rules for coding the articles have been clearly defined and strictly followed. Secondly, the chosen method, evaluative assertion analysis (EAA) (Osgood, Suci & Tannenbaum, 1957), represents one of the stricter varieties of content analysis. The use of EAA reduces the influence that the coder's personal opinions and experiences have on the coding¹⁰. Thirdly, a student has completed duplicate coding of a sample of articles with encouraging results: The difference in cumulative evaluative score amounted to a mere 5.31%.¹¹ These scores measure on a range from -9 to +9 how negatively or positively an author or authors of a comment views the ceasefire. Technically, the coder takes the product of the 'verb connector' and the 'evaluator' for each assertion and averages it across the commentary under examination.¹²

⁸ Viser (2003) notes that *Haaretz* and the *NYT* took a similar stance in the beginning of our examination period and that the *NYT* has more pronouncedly increased its pro-Israel partiality during the second Intifada and after the terrorist attacks of 9/11 than *Haaretz*.

⁹ When the ceasefire was announced on a Saturday, on which Israeli newspapers are not published, the following two issues (one and two days after the truce was announced) were considered.

¹⁰ A number of studies including Holsti (1962) seminal examination of John Foster Dulles' attitudes towards the Soviet Union have used the approach. Bos & Tarnai (1999), in their overview of the development of empirical content analysis, discuss the major criticisms of EAA

¹¹ The small number of cases and coders prevented us from employing other evaluative criteria such as Cohen's Kappa and Krippendorff's Alpha.

¹² For example, the sentence 'The agreement will (2) be broken (-3)' is coded -6, the sentence 'The tenant plan for a ceasefire agreement is (3) unsupported by both parties (-3)' is coded -9 and the sentence 'the agreement will lead (2) to peace (3)' would have been coded 6. 'The ceasefire agreement/the Tenant agreement .../the agreement' refer to the ceasefire in question and not to a future or past one, and therefore they are included in the overall analysis. Please refer to the Online appendix for further examples and the coding guide.

The predictive power of the stock market: In order to measure the predictive power of the stock market, the event study method was utilised. Event studies examine the returns to financial assets, i.e. the changes in price of a particular asset between one period and the next. Such studies attempt to measure investor reactions to specific incidences which are of particular interest to the researcher. Chen & Siems (2004), for example, apply the technique to assess the effects of terrorism on global capital markets.

In order to perform an event study, the timeframe of analysis is divided into two periods, the estimation and event windows. Financial data from the estimation window are used to assess the ‘normal’ returns of the asset of interest. These normal returns are then compared to the returns occurring in the event window, in which the incident of interest occurs. Deviations from the normal returns, predicted according to the estimation window, are called ‘abnormal returns’ and can be interpreted as the result of the event that is, in the case of this study, the announcement of a ceasefire. The event window will include the day of the announcement, plus or minus a few days. We will apply an event window of $[-3, 0]$ ¹³. The estimation window chosen for the purposes of this study was fifty days before the event up until ten days before the event $[-50, -10]$. Typical event studies utilise continuous periods for an estimation window often of up to 100 weeks before the occurrence of an event (Guidolin & La Ferrara, 2007). In addition, it allows for a satisfactorily large estimation window of 40 days, without leading to any difficulties with significantly overlapping estimation and event windows.

To measure normal performance we use the market model which assumes a linear relationship between the return of a given security and the return of the market portfolio in which the proportion invested in each security corresponds to its relative market value.

Formally expressed, for any asset i , R_{it} and R_{mt} are the returns on the asset during period t and the market portfolio respectively:

$$R_{it} = \alpha_i + \beta_i R_{mt} + \varepsilon_{it} \quad (1)$$

where ε_{it} is the disturbance term with a mean of zero and α_i and β_i are parameters. By applying the statistical model, one can calculate the market model parameters for the

¹³ The Online appendix also reports the results for a different event window $[-5, 0]$ as to ensure that the results are robust to the size of the event window.

estimation window using an ordinary least squares regression. Using the market parameters generated in this stage and the returns on the market portfolio for the event window, we are able to predict the expected normal returns for the security of interest during the event window.

The next step in the event study process is to measure the abnormal returns to the security of interest over the event window. This is, once again, done by using the market model parameter estimators. We can measure abnormal returns to security i for event time τ , $AR_{i\tau}$, within the market model as:

$$AR_{i\tau} = R_{i\tau} - \alpha_i - \beta_i R_{m\tau} \quad (2)$$

where α_i and β_i are the previously estimated market model parameters for security i , and $R_{m\tau}$ are the returns to the market portfolio over the event window τ . Finally, the abnormal returns are aggregated over the event window in order to accommodate the multi-day event windows. These cumulative abnormal returns (CARs) then function as the independent variable in the second stage of the analysis and are used to predict the success of the ceasefire agreement.

In order to employ the statistical model, it is necessary to select a stock that should be strongly affected by the announcement of a ceasefire. Furthermore, a market portfolio, which is typically a broad based stock index, is also necessary for the application of the market model. Tourism securities were chosen predicated on the assumption that as they mainly provide services in Israel, they should be more sensitive to developments in local conflict. Moreover, defence stocks were selected as their products were mostly designed for military use. We report the two results obtained for the one asset in each sector which provided the strongest results.¹⁴ The market portfolio, used to control for market fluctuations, is the TA-100 which includes a broad portfolio of securities from a wide range of industries.

All financial data has been retrieved through *Thomson Datastream*. On a number of occasions, a ceasefire is announced on Friday or Saturday, when the TASE is closed. Past studies (e.g., Daly & Paytakhti Oskooe, 2015) have tried to circumvent this problem of events on non-trading days by interpolating returns in order to fill in missing values. However, this procedure is unlikely to adequately substitute for

¹⁴ We selected companies in the tourism (*Dan Hotels Corporations Ltd.*, *Millenium Holdings*), defence (*Imco Industries Ltd.*, *Rabintex*, *Tat Industries Ltd.*) and insurance sectors (*Clal Insurance Enterprise Holdings*, *Phoenix Holdings Ltd.*), which we assumed to be particularly sensitive to changes in security policy. Results obtained for stocks originated in the insurance sector led to inconclusive results that the consulted experts in this market could not explain.

missing return observations (Bechtel & Schneider, 2010). For this reason, the dates of ceasefire announcements on Friday or Saturday were artificially transferred to Sunday, when investors are first given the opportunity to trade on the basis of the latest ceasefire announcement.

Evaluating predictive accuracy: To evaluate predictions, O'Brien (2002, 2010, see also Brandt, Freeman & Schrodt, 2014), applies different forecasting performance metrics. *Accuracy* measures the ability of the model to correctly differentiate between those objects of interests that do or do not experience some specified event. Formally, it is the number of correct predictions divided by the number of predictions made. We differentiate here, in the denominator, between the real and the potential predictions made. *Recall*, on the other hand, refers to the ability of the model to predict the event of interest. It amounts to the number of correctly predicted successes divided by the number of successes. Finally, the *precision* score pertains to the ability of the model to make predictions concerning the occurrence of interest without producing too many false positives. It is the number of correctly predicted successes divided by the number of predicted successes. This final measure, in the context of the present research, will be used to summarise the ability of each of the two competing models to determine the anticipated outcome of the relevant ceasefire.

For the use of O'Brien's (2002, 2010) forecast performance criteria, it is necessary to classify both the dependent and the independent variables dichotomously. In order to categorize the outcome of a ceasefire as successful or failure, it is necessary to set an arbitrary threshold, above which the ceasefire is deemed to be a success. Two thresholds of 10 and 14 violent events following the announcement of the ceasefire will also be considered. Both the stock market model and the journalist model will produce predictions that are positive or negative values with an expected mean of zero. It is, therefore, logical to assume that positive values, in terms of either journalists' cumulative evaluative scores or cumulative abnormal returns (CARs) on securities of interest can be interpreted as optimistic predictions, whilst negative values can be seen as pessimistic predictions. The CARs on defence stocks will be reversed, i.e. multiplied by minus one, for this stage of the analysis.¹⁵

¹⁵ The Online appendix displays the similar results of Receiver Operating Characteristic (ROC) analyses.

Results

It is important to note that there are certain ceasefire agreements on which the newspapers did not report at all. This is the case for *Haaretz* in six of the 24 ceasefires and in three ceasefire events for the *Jerusalem Post*. A long history of military censorship and possibly self-censorship, even among leading Israeli media might, at least partly, explain this divergence (e.g. Rinnawi, 2007). Table I lists the 24 ceasefires to be analysed in the present study.

[Table I in here]

The list of ceasefires reflects only a partial overlap between the non-reported ceasefires for the two Israeli newspapers. This stresses that military censorship possibly accounts for the partial silence of the media. The *Jerusalem Post*, the most conservative daily, is the most pessimistic media outlet with a prediction of six successes and 15 failures, while the NTY anticipates 15 failures and nine successes.

Table II reports the tests of the first two hypotheses. If H1 finds support within the data, there should be a negative and significant relationship between the cumulative evaluative score (CES) and the number of violent events occurring in the 14 day period following a ceasefire for the three newspapers.¹⁶ The event study analysis provides a test of the second hypothesis with its two sub-hypotheses; namely that (a) cumulative abnormal returns (CARs) of tourism stocks should negatively correlate with violations of a ceasefire; and (b) defence CARs should positively correlate. We report here the results for a threshold level of 14 violent events during two weeks after the truce was announced, and for the two assets under examination, with an event windows spanning over three days. Alternative specifications with event windows covering five days for the two assets, a time horizon of seven days and a threshold of ten events are reported in the Online appendix.

[Table II in here]

¹⁶As the robust standard errors are smaller, most likely due to small sample bias (Angrist & Pischke, 2009: 307, 296), we report the 'normal' standard errors, adopting the rule of thumb to list the larger of the two errors.

What immediately become apparent in Table II are the Alpha values, which are significantly different from zero, indicating that the negative binomial model is superior to that of the Poisson model. The coefficients and the related incidence rate ratios are rather small. This means that both the stockbrokers and the journalists rather under- than overestimate the possible violence following a truce. The opinions of *Jerusalem Post* commentators correlate negatively and significantly with the number of violent events which occur in the following two weeks, while the impact of the equivalent assessments made by *Haaretz* and *NYT* journalists are not significant.

The results nevertheless support the first theoretical expectation, demonstrating an increase in the chance that the ceasefire will be successful the more optimistic the reaction of media commentators to the announcement of a ceasefire becomes. Furthermore, the CARs of *Dan Hotels* are statistically significant and in the expected (negative) direction. This provides moderate support for H2(a) which states that the stronger the rally of tourism stocks in anticipation of the announcement of a ceasefire is, the more likely the ceasefire to be a success. The correct, i.e. negative, signs displayed by the tourism coefficients support the concept that tourism securities should react negatively to anticipated ceasefire failures, while the influence of the defence stock is not significant. Hence, only the tourist company, *Dan Hotels*, but not the defence firm, *Imco Industries*, reacts strongly and in the expected direction to the ceasefire agreements.

Comparing predictive accuracy: Figure 1, which reports the performance metrics accuracy, recall and precision, offers some interesting findings with regards to the oracle versus crowd controversy. Note that Accuracy 1 refers to the number of correct predictions out of the 24 ceasefire events whereas Accuracy 2 refers to the number of correct predictions out of the number of predictions made by the specific predictor. Firstly, it is clear that *Dan Hotels* is the best predictor of the fate of ceasefires, in terms of accuracy 1 and recall. In fact, *Dan Hotels* predicts the fate of ceasefires with 67% accuracy using the threshold of 14 violent events and with 57% accuracy with a threshold of 10 violent events, on par with the *New York Times*. The recall score, which measures the rate of true positives, is 62% for the 14 event threshold, stronger than those of the *New York Times*. Once again, the threshold of the 10 events is equal to that of the US newspaper at 56%. In terms of precision, or rate of false positives,

Dan Hotels perform very well with the 14 event threshold, whilst being somewhat weaker using the 10 events threshold.

It is important to notice the accuracy, recall and precision scores produced by the *Jerusalem Post*. Whilst the accuracy and recall scores are weak, between 33% and 42%, the precision scores are very high, reaching 100% for the 14 event threshold. The conservative nature of this journal might account for the tendency among its commentators to underpredict successful ceasefire. Out of a total 21 predictions made using JP commentaries, only six of them predicted ceasefires to be successful. However, those six were correctly predicted. *Haaretz*, with only 18 predictions out of the 24 ceasefire events, also scores well in the precision category. The predictive ability of the *New York Times* commentators is quite dismal, although its precision score is as good as those of the two assets. Another point to notice is the weakness of the *Imco Industries*' predictions. None of the *Imco Industries* combinations predict the anticipated outcome of the ceasefire events with accuracy or recall scores greater than 50%, which is equal to that of a coin toss. Only in two cases does *Imco Industries* achieve a precision rate of 60%, a result which is hardly convincing.

[Figure 1 about here]

O'Brien (2002: 805) sets 55% rate of accuracy, recall and precision which he describes as 'minimally acceptable'. All of the performance metrics of *Dan Hotels* score above this threshold, whilst *Imco Industries* predictions manage to do so only once. The *New York Times* fails three times to meet this minimum threshold, and the two domestic newspapers- twice. In his later paper O'Brien (2010: 95) sets a higher threshold of 67%, which is met by *Dan Hotels* in two out of four performance metrics; but is only met once by the other 'forecasters'.

From the above analysis it is possible to draw two cautious conclusions. Firstly, *Dan Hotels* steadily performs better than *Imco Industries*, the *New York Times*, *Haaretz* and the *Jerusalem Post*, as a predictor of the fate of ceasefires. Secondly, *Imco Industries* performs at least as good as the newspapers. These conclusions provide conditional support for Hypothesis 3. Although the defence firm *Imco Industries* fails to clearly provide better predictions than the media experts, the tourism company does so convincingly. Therefore, it is possible to conclude that, in line with Hypothesis 3, we have found evidence that movement in the share price of

the tourism firm, *Dan Hotels*, is a better predictor of ceasefire success than media experts at the media outlets examined in this article.

The Brier score results largely confirm these results. For an estimation window of five days, *Dan Hotels* have a Brier score of 0.218; the corresponding figure for *Imco Industries* is 0.222. Interestingly, the *New York Times* forecasts, with a Brier value of 0.16, are slightly more accurate than the predictions derived from the event study analyses, while the two Israeli newspapers reach values of 0.249 (*Haaretz*) and 0.267 (JP).¹⁷

Conclusion

Our evaluation of the competing crowd versus oracle theses about the usefulness of using the news or the financial industry for the prediction of political events has two main implications. Firstly, the statistical results show that, through analysis of stock market reactions, policy makers are better able to judge the likely success of a ceasefire and can in turn create policy using this information. Whether the actions resulting from this insight are that, in light of a promising ceasefire, greater labours are made to rebuild conflict-damaged areas, or that, facing a weaker agreement, greater efforts are made to stabilise the peace- this knowledge can be beneficial. Furthermore, as the results of this article show, one of the most conventional methods of information gathering in policy-making circles, the reading of newspaper commentaries, is often inferior to the insight that can be gained from an analysis of the stock market.

Secondly, the results of this article have relevance for the academic world. The study offers evidence for the argument that semi-experts – the ‘crowd’ – can make better predictions through their valuation of politically-sensitive stocks than single experts or ‘oracles’ can. Third, the study contributes to the slowly growing body of political economy literature that utilises financial markets as a tool for examining the link between security policy making and the economy (Schneider & Troeger, 2006; Guidolin & La Ferrara, 2007; Bechtel & Schneider, 2010). Fourth, this article

¹⁷ Brier scores measure the mean square error of the predictions, simultaneously addressing calibration (the alignment between forecasted probabilities and the binary outcomes) and sharpness (the ability to predict extreme probabilities close to the binary outcomes). Brier score vary between 0 and 2, with lower values indicating more accurate predictions. For an introduction to this and alternative measures in the context of conflict forecasts, see Brandt, Freeman & Schrodt (2014).

compares different types of forecasts and thus moves beyond the extant literature which mainly relies on the same source of information in the model evaluation.

We are aware that other types of experts might have performed better in predicting the accuracy of ceasefires than the media commentators selected for this study. This might be particularly the case for intelligence analysts (Mandel & Barnes 2014) and the ‘superforecasters’ identified by Mellers et al. (2014) and Tetlock & Gardner (2015). Our evaluation of whether or not different methods of prediction provide a better glimpse the future is therefore only an initial step for the competitive study of conflict forecasts which rely on judgemental information.

Replication data

The dataset, codebook and do-files for the empirical analysis in this article can be found at <https://www.prio.org/JPR/Datasets/> and <https://www.polver.uni-konstanz.de/gschneider/arbeitspapiere/replikationsdaten/>

Acknowledgments

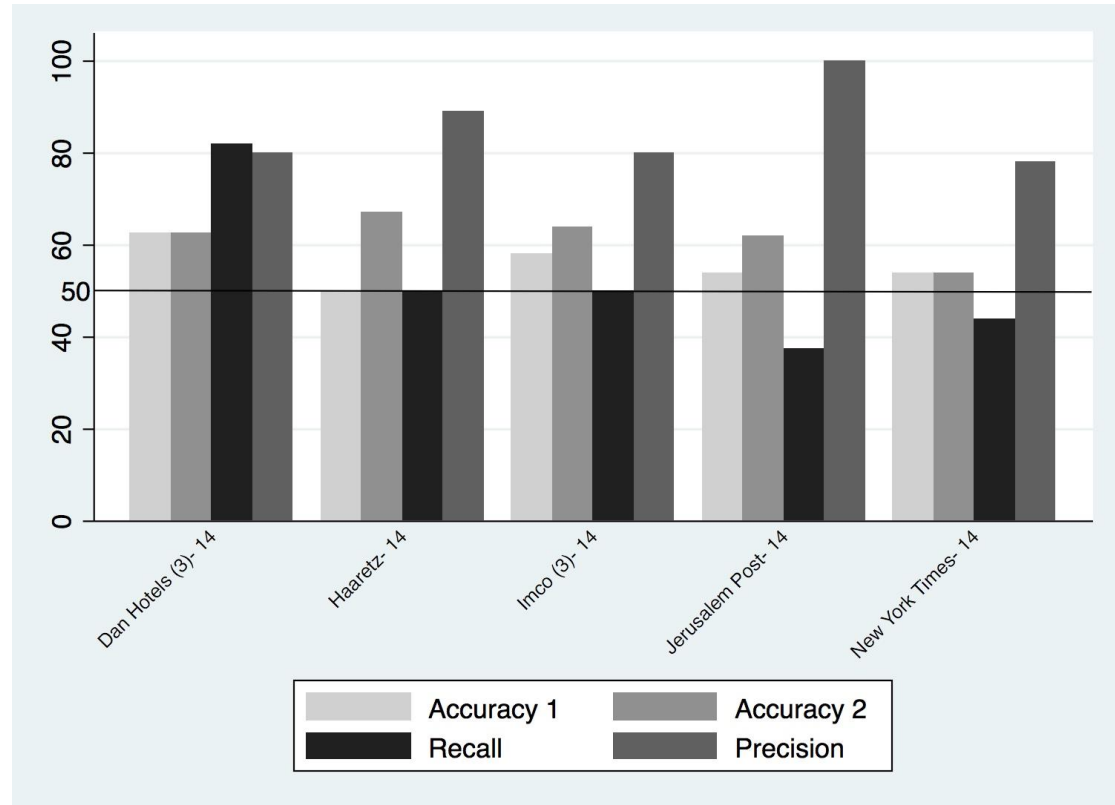
We would like to thank Adam Scharpf for research assistance at an early stage of the project and the special issue reviewers and editors for their helpful comments.

Funding

The second author wishes to acknowledge funding provided by the Ernst-Ludwig-Ehrlich-Foundation.

Figures and Tables

Figure 1: Predictive accuracy of financial market assets and newspaper reports



Note: Accuracy 1 is calculated as the number of correct predictions divided by 24 ceasefires while Accuracy 2 is calculated as the number of correct predictions divided by the number of overall predictions made by the specific media outlet. Recall is the ratio of the number of correctly predicted successes and the number of successes, and precision divides the number of correctly predicted successes with the number of predicted successes. The calculations refer to a time horizon of 14 days and a violence threshold of 14 events. The event window used with regard to the assets' analysis is (-3, 0).

Table I. Ceasefires and the predictions of the different ‘forecasters’

	<i>Date</i>	<i>Number of violent events</i>	<i>Success / failure</i>	<i>Predictions</i>				
				<i>‘Crowd’</i>		<i>‘Oracles’</i>		
				<i>Dan Hotels</i>	<i>IMCO</i>	<i>Haaretz</i>	<i>JP</i>	<i>NYT</i>
1	31 July 1993	9	S	S	F	S	S	S
2	05 March 1996	12	S	F	F	S	S	S
3	26 April 1996	7	S	F	F	S	F	F
4	30 September 2000	69	F	F	F	-	-	F
5	08 October 2000	51	F	F	F	-	F	S
6	17 October 2000	52	F	F	F	F	F	F
7	02 November 2000	53	F	S	S	F	F	F
8	23 May 2001	22	F	F	S	F	F	F
9	13 June 2001	15	F	S	F	F	F	F
10	18 September 2001	13	S	S	F	S	F	S
11	16 December 2001	9	S	F	F	-	F	S
12	21 February 2003	9	S	S	F	-	S	F
13	29 June 2003	5	S	S	S	F	S	S
14	18 January 2005	6	S	S	S	-	-	F
15	07 February 2005	3	S	S	S	S	S	S
16	25 September 2005	11	S	S	S	F	-	F
17	30 July 2006	52	F	F	F	-	F	S
18	12 August 2006	29	F	F	F	S	F	F
19	25 November 2006	7	S	F	S	F	F	F
20	17 June 2008	6	S	F	F	S	S	S
21	16 January 2009	12	S	F	S	S	F	F
22	13 March 2012	6	S	S	F	F	F	F
23	21 November 2012	10	S	F	S	F	F	F
24	26 August 2014	4	S	S	S	S	F	F
Predicted successes (S), failures (F):				11S, 13F	10S,14F	9S, 9F	6S,15F	9S,15F
Correctly predicted successes				9/16	8/16	8/16	6/16	7/16
Correctly predicted failures				6/8	6/8	4/8	7/8	6/8
Success is defined as the occurrence of up to 14 violent events within 14 days following the ceasefire declaration.								

Table II: The effects of ceasefire-induced Cumulative Abnormal Returns and newspaper ceasefire assessments on the number of subsequent violent events (14 days time frame, 14 violent events)

<i>DAN HOTELS</i>	0.21*** (0.01)				
<i>IMCO INDUSTRIES</i>		0.01 (0.02)			
<i>HAARETZ CES</i>			-0.12 (0.08)		
<i>JERUSALEM POST CES</i>				-0.332** (0.12)	
<i>NEW YORK TIMES CES</i>					-0.005 (0.05)
<i>Constant</i>	3.01*** (0.18)	2.98*** (0.18)	2.70*** (0.18)	2.53*** (0.19)	2.92*** (0.18)
<i>N</i>	24	24	18	21	24
<i>Log (pseudo) likelihood</i>	-94.43	-95.23	-65.44	-78.34	-94.92
<i>Alpha</i>	0.67 (0.19)	0.70 (0.20)	0.51 (0.18)	0.48 (0.16)	0.68 (0.20)

Note: Standard errors are given in parentheses. The larger standard error, be it robust or conventional is reported. In the negative binomial models this is generally the conventional standard error. * $p \leq 0.1$; ** $p \leq 0.05$; *** $p \leq 0.01$ using a one tailed test.

References

- Angrist, Josuha D & Jörn-Steffen Pischke (2009) *Mostly harmless econometrics: An empiricist's companion*. Princeton: Princeton University Press.
- Arrow, Kenneth J; Robert Forsythe; Michael Gorham; Robert Hahn; Robin Hanson; John O Ledyard; Robert Litan; Paul Milgrom; Forrest D Nelson; George R Neumann; Marco Ottaviani; Thomas C Schelling; Robert J Shiller; Vernon L Smith; Erik Snowberg; Cass R Sunstein; Paul C Tetlock; Philip E Tetlock; Hal R Varian; Justin Wolfers & Eric Zitzewitz (2008) The promise of prediction markets. *Science* 320(5878): 877.
- Bechtel, Michael M & Gerald Schneider (2010) Eliciting substance from 'hot air': Financial market responses to EU summit decisions on European defense. *International Organization* 64(2): 199-223.
- Berlin, Isaiah (1957) *The hedgehog and the fox. An essay on Tolstoy's view of history*. New York: New American Library.
- Berrebi, Claude & Esteban F Klor (2005) The impact of terrorism across industries: An empirical study. *CEPR Discussion Papers* 5360, C.E.P.R. Discussion Papers.
- Bos, Wilfried & Christian Tarnai. (1999) Content analysis in empirical social research. *International Journal of Educational Research* 31(8): 659 – 671.
- Brandt, Patrick T; John R Freeman & Philip A Schrodtt (2011) Real time: time series forecasting of inter- and intrastate political conflict. *Conflict Management and Peace Science* 28(1): 40–63.
- Brandt, Patrick T; John R Freeman & Philip A Schrodtt (2014) Evaluating forecasts of political conflict dynamics. *International Journal of Forecasting* 30(4): 944-962.
- Bueno de Mesquita, Bruce (1990) Pride of place: The origins of German hegemony. *World Politics* 43(1): 28-52.
- Bueno de Mesquita, Bruce (2009) *The predictioneer's game: Using the logic of brazen self-interest to see and shape the future*. New York: Random House.
- Bueno de Mesquita, Bruce (2010) Judging judgment. *Critical Review* 22(4): 355-388
- Bueno de Mesquita, Bruce (2011) A new model for predicting policy choices: Preliminary Tests. *Conflict Management and Peace Science* 28(1): 64-84.
- Brune, Amelie; Thorsten Hens; Marc Oliver Rieger; & Mei Wang (2015) The war puzzle: Contradictory effects of international conflicts on stock markets. *International Review of Economics*. 62(1): 1-21.
- Camyar, Isa & Bahar Ulupinar (2015) War and the sectoral distribution of wealth: Evidence from United States firms. *International Studies Quarterly* 59(4): 641–653.
- Chadefaux, Thomas (2014) Early warning signals for war in the news. *Journal of Peace Research* 51(1): 5–18.
- Chen, Andrew H. & Thomas F Siems (2004) The effects of terrorism on global capital markets. *European Journal of Political Economy* 20(2): 349 – 366.
- Daly, Vince & Paytakhti Oskooe, Seyyed Ali (2015) Stock market efficiency in Iran: Unit root testing with smooth structural breaks and non-trading days. *Economic Discussion Papers* 2015-6, Kingston University.
- de Caritat, Marie Jean Antoine Nicolas (Marquis de Condorcet) (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Paris: Imprimerie royale. . [Essay on the application of analysis to the probability of decisions rendered by a plurality of votes].

- Fama, Eugene F (1970) Efficient capital markets: A review of theory and empirical work. *Journal of Finance* 25(2): 383–417.
- Fleischer, Aliza & Steven Buccola (2002) War, terror, and the tourism market in Israel. *Applied Economics* 34(11): 1335–1343.
- Galton, Francis (1907) Vox populi. *Nature* 75:7.
- Guidolin, Massimo & Eliana La Ferrara (2007) Diamonds are forever, wars are not: is conflict bad for private firms? *American Economic Review* 97 (5): 1978–1993.
- Holsti, Ole R (1962) The belief system and national images: A case study. *Journal of Conflict Resolution* 6(3): 244–252.
- Le Bon, Gustave (1895) *Psychologie des foules*. Édition Félix Alcan, Paris [Psychology of Fools].
- List, Christian & Robert E Goodin (2001) Epistemic democracy: Generalizing the Condorcet jury theorem. *Journal of Political Philosophy* 9(3), 277-306.
- Mandel, David R & Alan Barnes (2014) Accuracy of forecasts in strategic intelligence. *PNAS* 111(30): 10984–10989.
- Meehl, Paul E. (1954) *Clinical versus statistical Prediction. A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.
- Mellers, Barbara; Lyle Ungar; Jonathan Baron; Jaime Ramos; Burcu Gurcay; Katrina Fincher; Sydney E. Scott; Don Moore; Pavel Atanasov; Samuel A. Swift; Terry Murray; Eric Stone; Philip E. Tetlock (2014) Psychological strategies for winning a geopolitical forecasting tournament. *Psychological Science* 25(5): 1106-1115
- Neumayer, Eric (2004) The impact of political violence on tourism: dynamic cross-national estimation. *Journal of Conflict Resolution* 48(2): 259.
- Nofer, Michael & Oliver Hinz (2014) Are crowds on the internet wiser than experts? The case of a stock prediction community. *Journal of Business Economics* 84(3): 303–338.
- O'Brien, Sean P (2002) Anticipating the good, the bad, and the ugly: An early warning approach to conflict and instability analysis. *Journal of Conflict Resolution* 46(6): 791-811.
- O'Brien, Sean P (2010) Crisis early warning and decision support: Contemporary approaches and thoughts on future research. *International Studies Review* 12(1): 87–104.
- Osgood, Charles E; George Suci & Percy Tannenbaum (1957) *The measurement of meaning*. Urbana: University of Illinois Press.
- Rinnawi, Khalil (2007) De-legitimization of media mechanisms: Israeli press coverage of the Al Aqsa Intifada. *International Communication Gazette* 69(2): 149-178.
- Rost, Nicolas; Gerald Schneider & Johannes Kleibl (2009) A global risk assessment model for civil wars. *Social Science Research* 38(4): 921–933.
- Schneider, Gerald (2014) Forecasting political developments with the help of financial markets. In *Predicting the Future in Science, Economics and Politics*, Frank Whelon Wayman, Paul R. Williamson, Bruce Bueno de Mesquita, Solomon Polacheck eds. Cheltenham: Edward Elgar, 213-231.
- Schneider, Gerald & Vera E Troeger (2006) War and the world economy: Stock market reactions to international conflicts, 1990-2000. *Journal of Conflict Resolution* 50(5): 623–645.
- Schneider, Gerald; Daniel Finke & Stefanie Bailer (2010) Bargaining power in the European Union. An evaluation of competing game-theoretic models. *Political Studies* 58(1): 85-103.

- Schneider, Gerald; Nils Petter Gleditsch & Sabine C Carey (2011) Forecasting in International Relations: One quest, three approaches. *Conflict Management and Peace Science* 28(1): 5–14.
- Sherden, William A (1997). *The fortune sellers: The big business of buying and selling predictions*. Hoboken, NJ: Wiley.
- Surowiecki, James (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*. Boston: Little, Brown.
- Taleb, Nassim N (2007) *The black swan: The impact of the highly improbable*. New York: Random House.
- Tetlock, Philip E (2005) *Expert political judgment: How good is it? How can we know?* Princeton University Press, Princeton.
- Tetlock, Philip E (2010) Second thoughts about expert political judgement: Reply to the symposium. *Critical Review* 22(4): 467-488.
- Tetlock, Philip E & Dan Gardner (2015) *Superforecasting. The art and science of prediction*. London: Random House.
- Thompson, Robert; Frans N Stokman; Christopher H Achen & Thomas König (2006) *The European Union Decides*. Cambridge: Cambridge University Press.
- Viser, Matt (2003) Attempted objectivity: An analysis of the New York Times and Ha'aretz and their portrayals of the Palestinian-Israeli conflict. *The International Journal of Press/Politics* 8(4): 114-120.

GERALD SCHNEIDER, b. 1962, Dr. phil. (University of Zürich, 1991); Professor of Political Science, International Relations Chair, University of Konstanz (1997–); Executive Editor of *European Union Politics* and co-Editor *International Interactions*.

MAYA HADAR, b. 1985, M.A. in Peace & Conflict (University of Haifa, 2013); Doctoral student at the Graduate School of Decision Sciences, University of Konstanz (2014-).

NAOMI BOSLER, M.A. in Politics and Management (University of Konstanz, 2010).