Meta-Induction and the Wisdom of Crowds: A Comment[*]

Christian J. Feldbacher-Escamilla

Winter 2012

Abstract

[367] In their paper on the influence of meta-induction to the wisdom of the crowd, Paul D. Thorn and Gerhard Schurz argue that adding meta-inductive methods to a group influences the group positively, whereas replacing independend methods of a group with meta-inductive ones may have a negative impact. The first fact is due to an improvement of average ability of a group, the second fact is due to an impairment of average diversity within a group by meta-induction. In this paper some critical remarks to meta-inductive group expansion and replacement are made. In particular it is stressed that both ability and diversity are of equal importance to a group's performance.

1 Introduction

In recent papers one of the authors of the article at hand (Thorn and Schurz 2012) has shown that some meta-inductive methods are optimal compared to competing methods, inasmuch as they are in the long run the most successful methods in a prediction setting (cf. especially Schurz 2008). Meta-inductive methods build their predictions on competing methods, depending on their past success. Since they depend on other methods, they normally decrease the diversity (or independence) within a setting. However, some very important results of social epistemology show that diversity in a setting is highly relevant for the whole performance within the setting. This is the so-called *influence of diversity on the wisdom of a crowd*. So, at first glance it seems that meta-inductive methods are valuable for their own sake only, but not for the sake of a whole group of methods' performance. For this reason Thorn and Schurz investigate the influence of meta-inductive methods on the performance of a group in

[*][This text is published under the following bibliographical data: Feldbacher-Escamilla, Christian J. (2012). "Meta-Induction and the Wisdom of Crowds. A Comment". In: *Analyse und Kritik* 34.2, pp. 367–382. DOI: 10.1515/auk-2012-0213. All page numbers of the published text are in square brackets. The final publication is available at https://www.degruyter.com/auk. For more information about the underlying project, please have a look at http://cjf.escamilla.academia.name.]

more detail. Since there are no general results about this influence in a broad setting, they perform simulations for quite specific settings. The main result of their argumentation and simulations is that

> "it is not generally recommendable to replace independent strategies by meta-inductive ones, but only to enrich them." (cf. Thorn and Schurz 2012, p.346)

[368] In this paper I am going to provide a complementary summary of and a critical comment about Thorn and Schurz's discussion. Complementarily to the authors' investigation I will introduce the basic concepts of the meta-inductive framework first (section 2). Afterwards the—by the authors discussed—new problem of how meta-inductive methods influence a group's performance will be presented in detail (section 3). In contrast to the author's result stated above, I want to highlight that adding meta-inductive methods to a group may also influence the group's performance negatively, although averaging the influence in a series of predictions supports Thorn and Schurz's recommendation, namely to enrich independent methods by meta-inductive ones (section 3.1).

Although Thorn and Schurz think that there is no general recommendation for replacing methods in a setting by meta-inductive ones, one could think that there is some kind of reliable heuristics favouring such a replacement. An argument along this line runs as follows:

1. Besides diversity, also average competence is influential to the whole performance within a setting. (results of social epistemology)
2. Meta-inductive methods normally increase the average competence on cost of diversity within a setting. (results of meta-induction)
3. Average competence is more influential to the whole performance within a setting than diversity is. (assumption)
4. Hence, meta-inductive methods normally improve the whole performance within a setting. (1–3)

Whether this argument is strong or not depends on whether average competence and diversity within a setting are the only influential factors for the group's performance (ad 1) and whether the increase of average competence within a setting by meta-inductive methods normally outweighs the loss of diversity in their influence on the group's performance (ad 2 and 3). As we will see in our discussion of specific cases of a group's performance, the first point holds for these cases. The second point is a question of choosing the right simulations: There are settings where average competence seems to be more influential to a group's performance than diversity and there are settings where diversity seems to be more influential than average competence. Whether or not meta-inductive replacement of independent methods is favourable depends on the situation under investigation. In a more or less critical addition to Thorn and Schurz's investigation I will argue by intuitions about group performances that generally both average competence and diversity are equally influential to the group's performance (section 3.2).

A terminological note ahead: It is common to identify methods with the agents that perform the methods. In our investigation I am following this convention, and so I will sometimes speak of agents, where I would have to speak of methods of agents and vice versa. I should also note that in the following I am mainly speaking about predictions. But the discussion and the results were most of the time equally correct or incorrect if I was speaking about decisions [369] or estimations. So in the following investigation the expressions 'prediction', 'decision' and 'estimation' are mostly interchangeable.

2 Meta-Induction and the Problem of Induction

One may distinguish two types of methods for performing actions within and by a group: (a) object-based methods and (b) meta methods (cf. a similar distinction in Schurz 2009, pp.200f). Object-based methods are only about the object the action is concerned with, whereas meta methods are also concerned with at least one method. Take, e.g., the action of predicting the value of a stock at a specific point in time. An object-based method for estimating the value would be concerned, e.g., with the stock, the stock market, the economy etc. only. In contrast, a meta method would be concerned, e.g., with the prediction method of a competing trader. So, if I, for example, try to predict the value of a stock at a certain time by studying the economic circumstances, a company's policy, the past values of the stock etc., then I am performing an object-based method. But if I just predict the value of the stock by copying the prediction of a competitor, I am using a meta method.

Each class of methods contains a specific subclass of inductive methods. Object-based inductive methods are like object-based methods, but in addition, their prediction of a value of a present or future event depends on the values of similar events of the past. In the case of meta-inductive methods, one's method about a present or future event is based on methods about events of the past (meta-inductive methods are called 'meta', because they are about methods and they are called 'inductive', because the methods they are about are methods based on events of the past). If I, e.g., have observed that there is a very competent trader whose method for predicting the right stock value worked very well in the past and if I copy her past event's value estimating method for predicting the future stock value, in doing so I am performing a meta-inductive method.

Meta-inductive methods are a relatively new way of considering and accounting for the classical problem of induction, namely the problem of how to justify at least some inductive methods. Classical approaches to this problem are often troubled by the fact that a justification of inductive methods by deductive methods alone is impossible and that a justification by inductive ones is circular or incomplete. But if one accepts as the only constraint for the justification of a method that it is to be shown one of the best of all performed methods in a situation (this is the so-called *best alternative approach* which seems to be the best alternative to the impossible justification *per se*), then one of the authors has given a very impressive justification of some meta-inductive methods for some very general situations (cf. the optimality results in Schurz 2008). This is done especially by Schurz in showing that in very general situations performing different kinds of meta-inductive methods allows one to reach an optimum in attaining a target such as, e.g., being most accurate in one's predictions etc.

Up to this point I have given only a classification of methods without clarifying the underlying notion, which will be done now: Very generally speaking, [370] methods allow one to achieve a target from a specific starting point with some instructions. Take, for example, the method of falsification. Here the starting point is a theory T and an observational sentence S. Intended with an application of the method is a falsification of T by S. Instructions are: 'Try to derive S from T and check with an experiment whether S turns out to be false!' and the like. Such a view on methods is technically implemented in the meta-inductive framework via functions. Instructions are represented by the definition of a function, starting points are represented by the arguments, targets are represented by the values of a function. One could, e.g., define a falsificationist function that maps theories T and observational sentences S to 0 (representing: 'not falsified by') and 1 (representing: 'falsified by'), depending on the derivability of S from T and the falsity of S. I have stated that for meta-inductive methods some methods of events of the past are relevant for similar events of the present or future. As usual, I will skip a discussion of the conditions for a similarity relation between events. But one may think of the throwing of a perfect die, or of the values of a stock at different times as paradigmatic examples of similar events. A sequence of similar events is indicated in the underlying framework via x (the starting event, e.g., a stock value at a specific point in time) and $x+1, x+2, \ldots, x+t$ (some following stock values at following time points). An object-based method for predicting the stock value at a specific time would be, e.g., a function from x to some value (points). An object-based inductive method for predicting the value at this time would be a function from $x, x-1, \ldots$ to some value. A meta method for predicting the value at this time would be, e.g., a function from a function from x to some value. And a meta-inductive method for predicting the value at this time would be a function from a function from $x, x-1, \ldots$ to some value.

Now, as already said, if the constraint for the justification of a method is its optimality in a situation, then, in the functional way of interpreting methods, to claim that a meta-inductive method is optimal in a situation or setting, is to claim that it is, e.g., one of the most accurate functions in the situation or setting. Figure 1 illustrates the optimality with respect to accuracy of a so-called weighted meta-inductive method ($V_{\alpha_{wMI}}$) that just takes the average of the predictions of the optimistic object-based method V_{α_1} of agent α_1 and the pessimistic object-based method V_{α_2} of agent α_2 in the given setting, where $V_{\alpha_T}(x)$ is taken to be the true value of the event x, that is the value that turns out to be the stock value at a specific point in time. The weighted meta-inductive method is defined here as:

$$V_{\alpha_{wMI}}(x) = c_{\alpha_1} \cdot V_{\alpha_1}(x) + c_{\alpha_2} \cdot V_{\alpha_2}(x)$$
$$\text{where (at the beginning) } c_{\alpha_1} \approx c_{\alpha_2} \approx 0.5 \tag{1}$$

which means that the optimistic and the pessimistic view are equally weighted (at the beginning). The factors $c_{\alpha_1}, c_{\alpha_2}, \ldots$ are called the 'weighting coefficients' and always sum up to 1.0. Their value depends on the past success of the predictors α_1 and α_2 which is measured by their past predictions ($V_{\alpha_1}(x-1)$ etc.) and the past values ($V_{\alpha_T}(x-1)$ etc.; an exact definition for the weighting coefficients is provided in equation 12). So, strictly speaking, the factors are also [371] functions of x, t and the agents $\alpha_1, \ldots, \alpha_n$ as we will see in section 3. If there were also a perfect predictor α_3 in the setting, i.e. $V_{\alpha_3} = V_{\alpha_T}$, then α_{wMI} would recognize this at some point in time (since the weighting coefficients are directly dependent on the past success of a predictor, the more successful a predictor is, the higher is the weighting coefficient for this predictor) and would weight the prediction of α_3 fully by completely neglecting the predictions of α_1 and α_2 in the long run:

$$V_{\alpha_{wMI}}(x) = c_{\alpha_1} \cdot V_{\alpha_1}(x) + c_{\alpha_2} \cdot V_{\alpha_2}(x) + c_{\alpha_3} \cdot V_{\alpha_3}(x),$$
$$\text{where } c_{\alpha_1} \approx c_{\alpha_2} \approx 0.0 \text{ and } c_{\alpha_3} \approx 1.0 \tag{2}$$

Figure 2 shows that α_{wMI} is optimal in this setting.

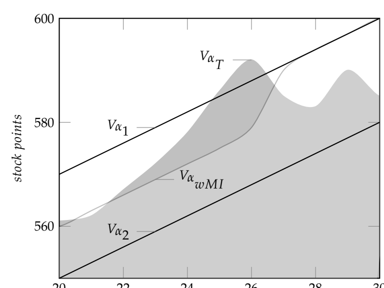

Figure 1: Setting: $\{\alpha_1, \alpha_2, \alpha_{wMI}\}$; since α_1 and α_2 are at the beginning equally near to the truth, α_{wMI} weights them equally. From day 23 on α_1's prediction is more accurate than that of α_2. Nevertheless it takes α_{wMI} three more days until it weights α_1's prediction higher than that of α_2, because until this time both competitors had lower success rates than α_{wMI}.

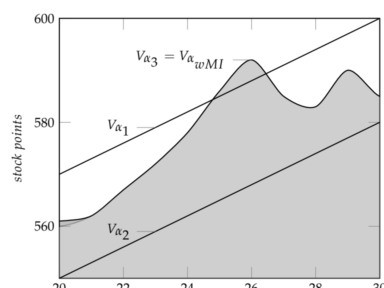

Figure 2: Setting: $\{\alpha_1, \alpha_2, \alpha_3, \alpha_{wMI}\}$; at day 20 α_{wMI} starts with the average of its competitor's prediction. Since from the beginning on only α_3's success rate is equal to or better than that of the meta-inductive method, α_{wMI} sticks also at the following days to the correct prediction.

(In both figures V_{α_T} are the stock points of AAPL *Apple Inc., NasdaqGS* (November 2012). V_{α_1} and V_{α_2} are feigned trend lines of the stock, using only preceding chart information of `http://www.nasdaq.com/symbol/aapl/`. All simulations in this paper were performed with scripts of the language *PERL*. A detailed description of this paper's simulations settings is to be found in the appendix.)

One important part of the meta-inductive research programme is to find meta-inductive methods that are optimal in different situations or to specify the conditions for situations where some well-known meta-inductive methods, such as, e.g., the weighted meta-inductive method, are optimal (I will give a general characterization of meta-inductive optimality results in section 3). As already mentioned, some optimality results in the framework of meta-induction seem to be very promising for accounting for the problem of induction. This is due to the step onto a meta level. Of course, by such a strategy the problem of how to justify object-based inductive methods still remains, but it can be shown that there are at least some inductive methods (namely some meta-inductive methods) that are justifiable in the sense that they are optimal. But now, since the [372] basic notions of the meta-inductive framework are introduced, let's come to the problem of how meta-induction influences the wisdom of the crowd.

3 Meta-Induction and the Wisdom of the Crowd

A detailled description of the wisdom of the crowd effect can be provided best by adding to the meta-inductive framework some equations of (Krogh and Vedelsby 1995): Take the group's prediction of the value of an event x to be the average of the individuals' decisions (cf. Krogh and Vedelsby 1995, p.232):

$$V_{\{\alpha_1,\ldots,\alpha_n\}}(x) = \frac{\sum_{i=1}^{n} V_{\alpha_i}(x)}{n} \tag{3}$$

Now, if we want to compare the group's prediction with that of the individuals, then we cannot do this directly since the individuals' predictions may be heterogeneous. But we can compare the group's prediction indirectly via the error of the prediction: We introduce a measure for the error of a prediction simply by measuring its difference from the true value and square it in order to achieve equal comparability of under- and overestimations. First, we introduce a measure for the error of an individual's prediction (cf. Krogh and Vedelsby 1995, p.232):

$$E_\alpha(x) = (V_{\alpha_T}(x) - V_\alpha(x))^2 \tag{4}$$

Then one can define a measure for the individuals' error just by calculating the average of the error of each individual (cf. Krogh and Vedelsby 1995, p.232):

$$E_{\varnothing\{\alpha_1,\ldots,\alpha_n\}}(x) = \frac{\sum_{i=1}^{n} E_{\alpha_i}(x)}{n} \tag{5}$$

And similar to the individual's error we measure the error of the group's prediction simply by measuring the difference of the true value and the predicted value (cf. Krogh and Vedelsby 1995, p.232):

$$E_{\{\alpha_1,\dots,\alpha_n\}}(x) = (V_{\alpha_T}(x) - V_{\{\alpha_1,\dots,\alpha_n\}}(x))^2 \tag{6}$$

One only needs to reformulate the equations to see that the following *The Crowd Beats the Average Law* holds:

Observation ((cf. Page 2007, p.209) and (cf. Krogh and Vedelsby 1995, p.233)).

$$E_{\{\alpha_1,\dots,\alpha_n\}}(x) \leq E_{\varnothing\{\alpha_1,\dots,\alpha_n\}}(x) \tag{7}$$

So, it can be shown that in general the error of a prediction of a group is equal to or smaller than the average error of the group's members, which is a very general positive feature of applying a meta method in predicting the value of an event x. One can observe furthermore that there are two important factors that influence [373] the group's error. Besides the influence on $E_{\{\alpha_1,\dots,\alpha_n\}}(x)$ by $E_{\varnothing\{\alpha_1,\dots,\alpha_n\}}(x)$, there is also some influence by the so-called factor of *diversity of the predictions* of the group's members. The diversity of an individual's prediction is measured by its distance from the average prediction. And the diversity within a whole group is measured by averaging the diversities of the individuals' predictions (cf. Krogh and Vedelsby 1995, p.232):

$$D_{\{\alpha_1,\dots,\alpha_n\}}(x) = \frac{\sum_{i=1}^{n}(V_{\alpha_i}(x) - V_{\{\alpha_1,\dots,\alpha_n\}}(x))^2}{n} \tag{8}$$

With the help of this measure one can show that the diversity within a group also influences the group's error. *The Diversity Prediction Theorem*:

Observation ((cf. Page 2007, p.208) and (cf. Krogh and Vedelsby 1995, p.232)).

$$E_{\{\alpha_1,\dots,\alpha_n\}}(x) = E_{\varnothing\{\alpha_1,\dots,\alpha_n\}}(x) - D_{\{\alpha_1,\dots,\alpha_n\}}(x) \tag{9}$$

So, one can say that, in general, it holds that the lower the average error or the higher the diversity within a group, the lower the error of the group's prediction. Note again that the method for building up the group's prediction is a meta method.

As Thorn and Schurz importantly stressed, performing a meta-inductive method may undermine the performance of another meta method, especially the performance of the wisdom of the crowd method. In the following parts of this section we will consider the authors' discussion of undermining a wisdom of the crowd effect by performing a meta-inductive strategy. But first I have to characterize the method under investigation in more detail: I have claimed that $V_{\alpha_{wMI}}$ is a meta-inductive method because it ends up with its prediction by calculating the past success of the other methods in the setting. The influence of the past success of a method in the setting at hand is coded in the weighting coefficients of equation 1 and 2. The weighting coefficient c_{α_i} for the prediction of an agent α_i increases with the success of α_i and decreases with its failings. So we can define a measure for the past success of a method just by summing up its individual errors of the past and inverting the result. Since we want to make a standardization (interval: $[0, 1]$), we stipulate that no prediction exceeds a value V_{max}. Then we can define the average predictive success of an agent α_i until time t of the value of an event x by (cf. for a more general form Thorn and Schurz 2012, p.341):

$$succ_{x,t}(\alpha) = \frac{\sum_{i=1}^{t} 1 - \frac{E_\alpha(x+i)}{{V_{max}}^2}}{t} \tag{10}$$

With the help of this notion one can define a measure for the attractiveness of an agent α_1 for another agent α_2 simply by measuring the relative success of the method (cf. Thorn and Schurz 2012, p.341):

$$attr_{\alpha_2,x,t}(\alpha_1) = max(\{0, succ_{x,t}(\alpha_1) - succ_{x,t}(\alpha_2)\}) \tag{11}$$

[374] It holds: The higher the relative success, the higher the attractiveness. Note that the degree of attractiveness of an agent for herself is 0.0. Just by relativizing the relative attractiveness of a method to the whole relative attractiveness of all methods, we end up with our weighting coefficients:

$$c_\alpha(\beta, x, t) = \frac{attr_{\beta,x,t}(\alpha)}{\sum_{i=1}^{n} attr_{\beta,x,t}(\alpha_i)} \tag{12}$$

(where all agents of the setting are $\alpha_1, \ldots, \alpha_n$)

The notion of success of a prediction method allows us now also to state a general form of an optimality result:

> α is optimal in its prediction of the value of x at time t in the setting $\alpha \in \{\alpha_1, \ldots, \alpha_n\}$ iff there is a function g such that $succ_{x,t}(\alpha) \geq max(\{succ_{x,t}(\alpha_1), \ldots, succ_{x,t}(\alpha_n)\}) - g(t, x, n)$ and for g it holds: $\lim_{t\to\infty} g(t, x, n) = 0$.

So, as claimed already in section 2, to show that a method in a setting is optimal is to show that it is one of the most accurate methods performed in the setting, at least in the long run. One very important optimality result is about the optimality of α_{wMI}:

Observation ((cf. Thorn and Schurz 2012, Theorem 2)). α_{wMI} is optimal in its predictions on any event in any setting $\{\alpha_1, \ldots, \alpha_n\}$ if the individual error functions E_{α_i} are convex $\forall i \leq n$. (Furthermore it holds that $g(t, x, n) = \sqrt{\frac{n}{t}}$.)

(A marginal note to this result: the condition of convexity in this result seems to suppose some kind of regularity of nature—not, like sometimes in accounts for justifying object-based inductive methods, in the nature of things in general, but in 'the nature of success'.) According to this result one is epistemically justified for performing $V_{\alpha_{wMI}}$ in a setting for predicting the value of an event x. Now, let the setting be one in which some epistemic agents $\alpha_1, \ldots, \alpha_n$ have a disagreement about the value of x. Investigations of social epistemology suggest that, one good reason for the agents to perform a difference-splitting strategy, i.e. for updating their predictions to $V^*_{\alpha_1}(x) = \cdots = V^*_{\alpha_n}(x) = \frac{V_{\alpha_1}(x)+\cdots+V_{\alpha_n}(x)}{n}$, is that they perhaps can make use of a wisdom of the crowd effect in the situation. But what, if one of the agents performs a meta-inductive method? What, if, e.g., $\alpha_1 = \alpha_{wMI}$? Since imitating or weighting the predictions of other agents may decrease the diversity within the setting, according to equation 9 the wisdom of the crowd effect within the group may also decrease. And this could undermine the performance of the difference-splitting strategy in the setting. More generally one may put the problem at hand as follows:

> Let Γ_1 be a group of agents. What changes within the group Γ_1, resulting in a group Γ_2 of agents, adhere or improve a wisdom of the crowd effect with respect to an event x: $E_{\Gamma_2}(x) \leq E_{\Gamma_1}(x)$?

[375] Of course, there are infinitely many relevant answers to this question. The easiest would be perhaps: just form $\Gamma_1 = \{\alpha_1, \ldots, \alpha_n\}$ to a group of perfect predictors $\Gamma_2 = \{\alpha_1 = \alpha_T, \ldots, \alpha_n = \alpha_T\}$ and the relation above is guaranteed. Such investigations are subsumed under the label 'institutional design' in social epistemology. But what concerns the authors and what matters here is the question of how to change Γ_1 by meta-inductive strategies. So we are concerned especially with questions of meta-inductive institutional design which can be twofold: adding meta-inductive agents or replacing agents by meta-inductive ones (removing meta-inductive agents is similar to adding them, but comparing the results in the opposite direction).

3.1 Adding meta-inductivists to groups

One observation by the authors is that adding meta-inductive agents to a group does not harm and may even increase the performance of the group, if the meta-inductive agent operates optimally in the situation (cf. Thorn and Schurz 2012, p.346):

Observation. Let Γ_1 be a group of agents $\alpha_1, \ldots, \alpha_n$ and $\Gamma_2 = \Gamma_1 \cup \{\alpha_{wMI}\}$. Then it holds that $E_{\Gamma_2}(x) \leq E_{\Gamma_1}(x)$ provided that $E_{\alpha_{wMI}}(x) \leq E_{\alpha_i}(x)\ \forall i \leq n$.

One may note that adding any agent that satisfies this optimality constraint to a group does not harm the wisdom of the crowd effect of the group. But the difference between any non-meta-inductive agent and α_{wMI} is that the optimality results for α_{wMI} show that in the long run she is most probable to operate optimally in a situation. But of course, although it is probable that α_{wMI} makes an optimal prediction, she still may act sometimes non-optimally and by this decrease the wisdom of the crowd effect of the group. Figure 3 shows a simple simulation of such a situation. The wisdom of the crowd effect is illustrated here via a meta method $V_{\alpha_{WC}}$ (this technical opportunity was also already mentioned by the authors). Note that this method is not meta-inductive, since the weighting coefficients of $V_{\alpha_{WC}}$ are constant (not depending on the past success of the agents).

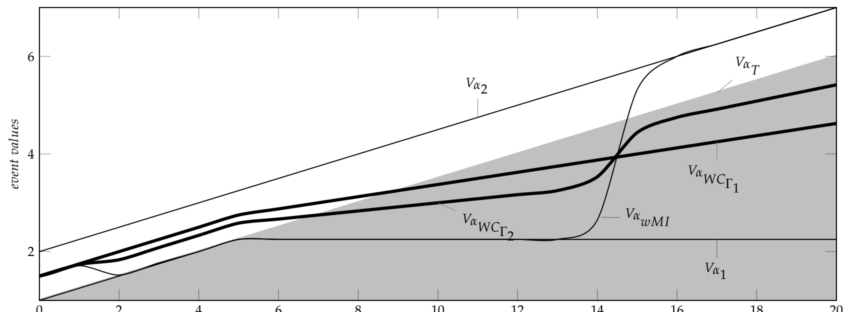

Figure 3: Setting: $\Gamma_1 = \{\alpha_1, \alpha_2\}$, $\Gamma_2 = \{\alpha_1, \alpha_2, \alpha_{wMI}\}$; as can be seen by considering the distances of $V_{\alpha_{WC_{\Gamma_1}}}$ and $V_{\alpha_{WC_{\Gamma_2}}}$ to the truth (V_{α_T}), adding the meta-inductivist $V_{\alpha_{wMI}}$ to a group normally increases (here: until $t < 8$ and starting again at $t > 14$) the wisdom of the crowd effect, but may also decrease it (here in the frame $8 \leq t \leq 14$). In the long run there will be more frames wherein the effect is unharmed than frames where it is harmed. Harmfull would be such an adding especially in an oscillating setting, where the accuracy of the predictions oscillates between α_1 and α_2. The following table lists the exact influence of adding $V_{\alpha_{wMI}}$ to a group in this setting, where $\Delta(t) = E_{\alpha_{WC_{\Gamma_1}}}(t) - E_{\alpha_{WC_{\Gamma_2}}}(t) = (V_{\alpha_T}(t) - V_{\alpha_{WC_{\Gamma_1}}}(t))^2 - (V_{\alpha_T}(t) - V_{\alpha_{WC_{\Gamma_2}}}(t))^2 = E_{\Gamma_1}(t) - E_{\Gamma_2}(t)$ (negative values indicate negative, positive values indicate positive influence):

t:	1	2	3	4	5	6	7	8	9	10
$\Delta(t)$:	0.013	0.139	0.139	0.139	0.139	0.113	0.063	−0.012	−0.111	−0.234
t:	11	12	13	14	15	16	17	18	19	20
$\Delta(t)$:	−0.382	−0.554	−0.750	−0.970	0.466	0.703	0.889	1.092	1.313	1.550

That adding an optimally operating meta-inductive agent α_{wMI} does not harm the wisdom of the crowd effect is due to the fact that it diminishes the average error of the individuals. So, although an optimally operating α_{wMI} diminishes the diversity within a group ($D_{\Gamma_2} \leq D_{\Gamma_1}$), this influence is always compensated by α_{wMI}'s also decreasing the average error in such a situation ($E_{\varnothing\Gamma_2} \leq E_{\varnothing\Gamma_1}$). A more general result which does not depend on the condition that the meta-inductive agent acts *de facto* optimally in a situation, can be provided not with respect to the wisdom of the crowd effect in single predictions, but with respect to averaging the wisdom of the crowd effects in multiple predictions:

Observation. Let Γ_1 be a group of agents $\alpha_1, \ldots, \alpha_n$ and $\Gamma_2 = \Gamma_1 \cup \{\alpha_{wMI}\}$. Furthermore, let us define the following average competence measures: [376]

- $E_{\alpha\varnothing t}(x) = \dfrac{\sum_{i=1}^{t} E_\alpha(x+i)}{t}$ is the measure for an agent α's average competence in predicting events of type x up to t.

- $E_{\varnothing\{\alpha_1,\ldots,\alpha_n\},t}(x) = \frac{\sum_{i=1}^{n} E_{\alpha_i \varnothing t}(x)}{n}$ is the measure for a group's average competence in predicting events of type x up to t.

- $E_{\Gamma,t}(x) = \frac{\sum_{i=1}^{t} E_{\Gamma}(x+i)}{t}$ is the measure for a group Γ's competence in predicting events of type x up to t.

Then it holds that $E_{\Gamma_2,t}(x) \leq E_{\varnothing\Gamma_1,t}(x)$ if $t \to \infty$.

So, adding a meta-inductive agent to a setting may decrease a wisdom of the crowd effect of the group, as simulated in figure 3, but averaging the wisdom of the crowd effects for multiple predictions shows that there is no harm in the long run.

3.2 Replacing agents by meta-inductivists

[377] In replacing independent agents by meta-inductivists, the simulations of the authors suggest the following heuristic view on meta-inductive institutional design:

- Transforming a group Γ_1 of independent agents to a group Γ_2 of only meta-inductive agents decreases a wisdom of the crowd effect with some random exceptions (cf. Thorn and Schurz 2012, tables 2–5).

- Transforming a group Γ_1 of, e.g., 10% experts (high competence, i.e.: low individual errors) and 90% non-experts that are not or nearly not incompetent (competence around 50%), to a group Γ_2 of only experts and meta-inductive agents increases a wisdom of the crowd effect (cf. Thorn and Schurz 2012, tables 7–9).

- In the case of local access (the so-called *Moore neighbourhood*), performing a meta-inductive strategy ends up with a better wisdom of the crowd effect than performing a difference-splitting strategy of a α_{WC} as local peer imitator (cf. Thorn and Schurz 2012, tables 10–12, 16–18).

- There is a—with respect to a wisdom of the crowd effect—better meta-inductive method than the above described α_{wMI}, namely a cautious weighted meta-inductive method. It results from adding a summand $g(t)$ to the success-rate of an opponent's method (equation 11 is changed to $attr_{\alpha_2,x,t}(\alpha_1) = max(\{0, succ_{x,t}(\alpha_1) + g(t) - succ_{x,t}(\alpha_2)\})$, where for g it holds: $\lim_{t\to\infty} g(t) = 0$). Colloquially speaking, one can say that it is easier to be attractive for a cautious weighted meta-inductive agent at the beginning of a series of predictions than at the end. This means that at the beginning of a series of predictions a cautious weighted meta-inductive agent shows more the pattern of an imitating agent, whereas at the end of such a series (when she increased her competence by imitation) she acts more meta-inductively selective. The cautious weighted meta-inductive agent performs quite well in an expert setting, i.e. in an initial group of highly competent and independent agents (cf. Thorn and Schurz 2012, tables 19–22).

Besides this heuristic, Thorn and Schurz claim that collective diversity (measured by $D_{\{\alpha_1,\ldots,\alpha_n\}}$) is not as important as individual ability (low error on average $E_{\varnothing\{\alpha_i,\ldots,\alpha_n\}}$) to the wisdom of a crowd:

> "we would also like to suggest, in contradiction to Page (2007, 208), that *collective diversity* [...] is not as important as *individual ability* [...] to the wisdom of a crowd [....]." (cf. Thorn and Schurz 2012, p.345)

They provide two reasons for their claim. Firstly, that it is practically seen relatively easy to increase the collective diversity within a group, whereas it is often impossible to decrease the individual errors. Secondly, that increasing individual ability by decreasing the average error is—independent of the influence [378] of diversity $D_{\{\alpha_1,\ldots,\alpha_n\}}$—sufficient for minimizing the group error $E_{\{\alpha_1,\ldots,\alpha_n\}}$, while increasing the collective diversity—independent of the influence of error on average $E_{\varnothing\{\alpha_1,\ldots,\alpha_n\}}$—is not. So, e.g., $E_{\{\alpha_1,\ldots,\alpha_n\}} = 0$ if $E_{\varnothing\{\alpha_1,\ldots,\alpha_n\}} = 0$, whereas it holds not generally $E_{\{\alpha_1,\ldots,\alpha_n\}} \to 0$ if $D_{\{\alpha_1,\ldots,\alpha_n\}} \to \infty$.

The first reason is a practical one and does not influence the relation between $E_{\{\alpha_1,\ldots,\alpha_n\}}$, $E_{\varnothing\{\alpha_1,\ldots,\alpha_n\}}$ and $D_{\{\alpha_1,\ldots,\alpha_n\}}$ as observed in equation 9. It seems to be relevant only in cases of evaluating different methods of decreasing the collective error of a group with similar results: Let us assume, e.g., that there are three groups Γ_1, Γ_2 and Γ_3 and let us assume further that Γ_2 and Γ_3 were formed by different methods of decreasing the collective error of Γ_1 to an equal level, i.e. $E_{\Gamma_2}(x) = E_{\Gamma_3}(x) < E_{\Gamma_1}(x)$. Say Γ_2 was formed by decreasing the individual errors $E_{\varnothing\Gamma_2}$ while keeping the group's diversity D_{Γ_2} equal, whereas Γ_3 was formed by increasing the group's diversity while keeping the individual errors of the group's members constant. One may claim that the method of transforming Γ_1 to Γ_2 is more preferable than that of transforming Γ_1 to Γ_3, since in many cases Γ_2 can be enhanced further to a group Γ_4 with $E_{\Gamma_4}(x) < E_{\Gamma_2}(x)$ by increasing collective diversity D_{Γ_4}. But, as $E_{\Gamma_2}(x)$ and $E_{\Gamma_3}(x)$ are equal, one cannot say that Γ_2 is wiser than Γ_3 or that the redesign of Γ_1 to Γ_2 is more preferable than the redesign of Γ_1 to Γ_3 with respect to their increasing the wisdom of the crowd effect only. Additionally one may also doubt that it is practically seen easier to increase diversity than to increase individual ability. Of course, one can always increase diversity by making bold over- or underestimations. But finding really adequate methods and algorithms for implementing smart diversity enhancing agents into a setting is a very difficult and subtle problem of machine learning theory (cf. Cunningham 2007).

The second reason is a theoretical one and concerns directly equation 9. In the following I am going to argue that this claim is oversimplified. Since meta-inductive methods normally decrease the average error of a group at the cost of its diversity and since the claims of the authors are in favour of evaluating the first effect higher than the second, the author's balance seems to be in favour of α_{wMI}'s 'enterprise of invading into a group and eliminating hostile agents', to put it in MI6's terms of institutional design. What are the reasons that undermine this different balancing of both factors' influence?

Firstly, it is easy to see that in the border case of an absolute wise crowd Γ_1, i.e.: $E_{\Gamma_1}(x) = 0$, the influence of both factors to the crowd's wisdom is exactly equal: $E_{\varnothing\Gamma_1}(x) = D_{\Gamma_1}(x)$. So, in such a case it does not matter whether the crowd's wisdom is due to minimal error on average or due to broad enough diversity, compensating for any errors on average. But if it does not matter in this case, why should it matter in other cases? Why should redesigning a group by a method that decreases error on average at the cost of diversity be more preferable than a method that increases diversity at the cost of increasing error on average? Again, to put it in less technical and more colloquial words:

> "we can say that, if the ensemble members are more likely on average to be right, and when they are wrong they are wrong at different points, then their decisions by majority voting are more likely to be [379] right than that of individual members. But they must be more likely on average to be right [i.e. the competence condition or the factor of error on average] and when they are wrong they must be wrong in different ways [i.e. the independence condition or the factor of diversity]." (cf. Cunningham 2007, p.2)

Secondly, also from a methodological point of view the importance of diversity is stressed very often. Paul Feyerabend, e.g., brought into the classical discussion of the unity of science a diversity argument, claiming that progress in science is sometimes possible only via diversity in, or plurality of theories and methods (cf. Feyerabend 1993, p.21, p.107). Figure 4 and 5 illustrate two simple cases of increasing a group's performance by increasing diversity. Figure 6 illustrates the group performance of some kind of a meta-anti-inductive method that increases partly diversity on cost of competence. Of course, this simple simulation serves only as a toy model and such cases may be shown to be relatively seldom in a [380] detailed description of the successful history of science. But if one takes Feyerabend's argumentation and critique seriously (of course always keeping in mind his role as *advocatus diaboli*), then some perhaps seldom, but nevertheless very important parts of the history of science are estimated as being successful due to diversity or plurality at the cost of ability with respect to an old paradigm.

Thirdly, in fact methods that increase diversity at the cost of competence are often performed. Perhaps this can be seen best by consideration of interdisciplinary research: We all are familiar with the fact that one important criterion for getting funding for research is interdisciplinarity. Behind this criterion stands the hope that increased diversity allows one to end up with better results than just by forming a group out of very competent researchers that act in concert, but that are for this reason sometimes also wrong in similar ways. In order to achieve better results through diversity, often experts of one area of

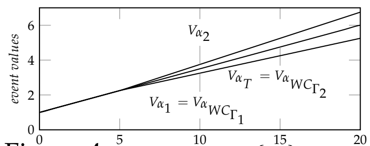

Figure 4: Setting: $\Gamma_1 = \{\alpha_1\}$, $\Gamma_2 = \{\alpha_1, \alpha_2\}$; In this redesign of Γ_1 to Γ_2 a wisdom of the crowd effect is increased to an optimum by introducing diversity on no cost of average ability.

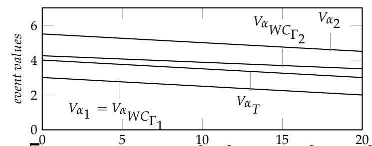

Figure 5: Setting: $\Gamma_1 = \{\alpha_1\}$, $\Gamma_2 = \{\alpha_1, \alpha_2\}$; In this redesign of Γ_1 to Γ_2 a wisdom of the crowd effect is increased by introducing diversity on cost of average ability.

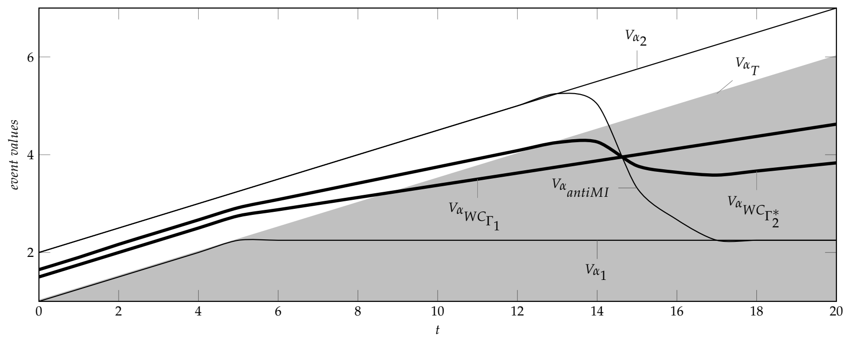

Figure 6: Setting: $\Gamma_1 = \{\alpha_1, \alpha_2\}$, $\Gamma_2^* = \{\alpha_1, \alpha_2, \alpha_{antiMI}\}$; by considering the distances of $V_{\alpha_{WC_{\Gamma_1}}}$ and $V_{\alpha_{WC_{\Gamma_2^*}}}$ to the truth (V_{α_T}), adding the meta-anti-inductivist $V_{\alpha_{antiMI}}$ to a group exceptionally increases (here: in the frame $10 < t < 15$) the wisdom of the crowd effect, but normaly decreases it. The attractivity of an agent's prediction increases with the agent's error rate, compared to the anti-meta-inductivists error rate (the error rate estimates similar to the success rate of an agent—cf. equation 10 and skip the inversion). The following table lists the exact influence of adding $V_{\alpha_{antiMI}}$ to a group in this setting (for Δ's definition cf. figure 3):

t:	1	2	3	4	5	6	7	8	9
$\Delta(t)$:	−0.194	−0.194	−0.194	−0.194	−0.194	−0.200	−0.188	−0.158	−0.111
t:	10	11	12	13	14	15	16	17	18
$\Delta(t)$:	−0.047	0.035	0.134	0.250	0.334	−0.391	−1.083	−1.778	−2.095
t:	19	20							
$\Delta(t)$:	−2.438	−2.804							

expertise enter an area they are not that familiar with and hence in which they are not that competent. But their hope is that such an enterance is fruitful to some extent. The assumption that diversity is of equal importance to wisdom of the crowd effects as competence is, would serve as a good explanation for the rationality of such a 'hope'. There are lots of other examples that point out the influence of diversity (cf. Surowiecki 2005) and (Page 2007). The assumption of such an influence has also become an essential part of a new strategy for approaches that try to diminish discrimination in different areas of life. Take, e.g., the new line of argumentation in feminism that stresses especially the influence of diversity on some wisdom of the crowd effects (cf., e.g., the summary in Fehr 2011, sect.7.2). If it can be shown that diversity in sex within a group also correlates with an increase of some wisdom of the crowd effects, than the influence of diversity seems to be a good justification for positive discrimination, even if, e.g., positive discrimination were at the cost of competence. (Note that recent studies suggest that it would turn out inversely: to assume a loss of competence by a vague criterion of being equally competent seems to be less plausible; more plausible seems also an increase of competence since women are shown to be unexpectedly often underestimated and so a vague criterion for being equally competent that may be in favour of women, would probably be compensated by such an underestimation (cf., e.g., Fehr 2011, sect.7.1)).

I think that such examples, discussed in more detail and with more empirical facts, could serve quite well for justifying the claim that diversity within a group is *de facto* of equal importance to wisdom of the crowd effects as competence is. But of course such a justification would always be only some kind of justification via abduction. And for abductive reasoning there are really little 'optimality results' at hand in the literature.

4 Summary & Conclusion

We have seen that the strategy of meta-induction to deal with classical epistemological problems not only on an object-based level, but also on a meta level brings some new insights into the discussion of these problems. In accordance [381] with the author's discussion we agree with the fruitfulness of adding meta-inductive methods into a setting, even with respect to the wisdom of the crowd where the factors *individual competence* and *diversity* are highly relevant. Nevertheless, it has to be stressed that for a positive influence in single predictions the performance of the single meta-inductivists matters a lot: the faster a meta-inductivist gets on the right track (i.e.: the function g of the optimality result for the meta-inductivist 'runs' faster against zero), the more positive is her influence on the wisdom of the crowd. Regarding a replacement of methods by meta-inductive ones, the author's simulations show that there is a meta-inductive method, namely the cautious weighted meta-inductivist, that behaves at least in a local accessible expert-setting quite well. Although they stress that there is no general recommendation for such a replacement, one may tend to weight the factors *individual competence* and *diversity* differently and so argue in favour of such a replacement. With the help of very general, but still realistic cases with positive group performance we have argued against such a different weighting. But whether meta-inductivistic replacement influences a group's performance positively or not depends of course on the detailed circumstances of the setting and has to be calculated and simulated case by case. The theoretical investigation of the authors provide a very good starting point for such simulations, because they give us some important hints about the best meta-inductivstic strategies to choose for meta-indictive institutional design.

References

Cunningham, Pádraig (2007-04). *Ensemble Techniques*. Tech. rep. UCD-CSI-2007-5. Dublin: UCD School of Computer Science and Informatics. URL: http://www.csi.ucd.ie/files/UCD-CSI-2007-5.pdf.

Fehr, Carla (2011). "What is in it for me? The benefits of diversity in scientific communities". In: *Feminist Epistemology and Philosophy of Science: Power in Knowledge*. Ed. by Grasswick, Heidi E. Dordrecht: Springer, pp. 133–156.

Feyerabend, Paul (1993). *Against method*. 3. ed. London: Verso.

Krogh, Anders and Vedelsby, Jespers (1995). "Neural Network Ensembles, Cross Validation, and Active Learning". In: *Advances in Neural Information Processing Systems 7*. Ed. by Tesauro, Gerald, Touretzky, David, and Leen, Todd. Cambridge: The MIT Press, pp. 231–238.

Page, Scott E. (2007). *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton: Princeton University Press.

Schurz, Gerhard (2008). "The Meta-Inductivist's Winning Strategy in the Prediction Game: A New Approach to Hume's Problem". In: *Philosophy of Science* 75.3, pp. 278–305. DOI: 10.1086/592550.

— (2009). "Meta-Induction and Social Epistemology: Computer Simulations of Prediction Games". In: *Episteme* 6.02, pp. 200–220. DOI: 10.3366/E1742360009000641.

Surowiecki, James (2005). *The Wisdom of Crowds*. New York: Anchor Books.

Thorn, Paul D. and Schurz, Gerhard (2012). "Meta-Induction and the Wisdom of Crowds". In: *Analyse und Kritik* 34.2, pp. 339–366.

Appendix

[382] Detailed description of the simulation settings (curves smoothed):

ad Figure 1: $V_{\alpha_1}(t) = 1 \cdot (t-20) + 570$; $V_{\alpha_2}(t) = 1 \cdot (t-20) + 550$; $V_{\alpha_T}(t) =$ stock points of AAPL *Apple Inc., NasdaqGS* (November 2012), available via the Nasdaq chart on AAPL at http://www.nasdaq.com/ symbol/aapl/; $c_{\alpha_1}(\alpha_{wMI}, x, t)$ and $c_{\alpha_2}(\alpha_{wMI}, x, t)$ are calculated via equations 10–12; Simulation of $V_{\alpha_{wMI}}$ (where x is intended to be interpreted as the event: development of AAPL): $V_{\alpha_{wMI}}(t+1) = c_{\alpha_1}(\alpha_{wMI}, x, t) \cdot V_{\alpha_1}(t+1) + c_{\alpha_2}(\alpha_{wMI}, x, t) \cdot V_{\alpha_2}(t+1)$.

ad Figure 2: To figure 1 similar setting with: $V_{\alpha_3}(t) = V_{\alpha_T}(t)$; $c_{\alpha_3}(\alpha_{wMI}, x, t)$ is also calculated via equations 10–12; Simulation of $V_{\alpha_{wMI}}$: $V_{\alpha_{wMI}}(t+1) = c_{\alpha_1}(\alpha_{wMI}, x, t) \cdot V_{\alpha_1}(t+1) + c_{\alpha_2}(\alpha_{wMI}, x, t) \cdot V_{\alpha_2}(t+1) + c_{\alpha_3}(\alpha_{wMI}, x, t) \cdot V_{\alpha_3}(t+1)$.

ad Figure 3: $V_{\alpha_1}(t) = min(\{2, \frac{2}{5} \cdot t\})$; $V_{\alpha_2}(t) = \frac{2}{5} \cdot t + 2$; $V_{\alpha_T}(t) = \frac{2}{5} \cdot t$; $c_{\alpha_1}(\alpha_{wMI}, x, t)$ and $c_{\alpha_2}(\alpha_{wMI}, x, t)$ are calculated via equations 10–12; Simulation of $V_{\alpha_{wMI}}$ (where x is an arbitrary event): $V_{\alpha_{wMI}}(t+1) = c_{\alpha_1}(\alpha_{wMI}, x, t) \cdot V_{\alpha_1}(t+1) + c_{\alpha_2}(\alpha_{wMI}, x, t) \cdot V_{\alpha_2}(t+1)$. Calculation of $V_{\alpha_{WCr_1}}$ and $V_{\alpha_{WCr_2}}$: $V_{\alpha_{WCr_1}}(t) = \frac{V_{\alpha_1}(t) + V_{\alpha_2}(t)}{2}$ and $V_{\alpha_{WCr_2}}(t) = \frac{V_{\alpha_1}(t) + V_{\alpha_2}(t) + V_{\alpha_{wMI}}(t)}{3}$.

ad Figure 6: To figure 3 similar setting with the exception: instead of α_{wMI} a meta-anti-inductivist α_{antiMI} is used, whose estimations are calculated via an attractivity measure for players which increases with past failings of a player: $V_{\alpha_{antiMI}}(t+1) = c^*_{\alpha_1}(\alpha_{antiMI}, x, t) \cdot V_{\alpha_1}(t+1) + c^*_{\alpha_2}(\alpha_{antiMI}, x, t) \cdot V_{\alpha_2}(t+1)$, where it holds (for the calulation of E_α cf. equation 4; $V_{max} = 3$; weighting at $t = 0$: $attr^*_{\alpha_{antiMI}, x, 0}(\alpha_2) = 1.0$):

$$attr^*_{\alpha_{antiMI}, x, t}(\alpha_1) = max(\{0, \frac{\sum_{i=1}^{t} \frac{E_{\alpha_1}(x+i)}{V_{max}^2}}{t} - \frac{\sum_{i=1}^{t} \frac{E_{\alpha_{antiMI}}(x+i)}{V_{max}^2}}{t}\})$$

similar calculation of $attr^*_{\alpha_{antiMI}, x, t}(\alpha_2)$

$$c^*_{\alpha_1}(\alpha_{antiMI}, x, t) = \frac{attr^*_{\alpha_{antiMI}, x, t}(\alpha_1)}{attr^*_{\alpha_{antiMI}, x, t}(\alpha_1) + attr^*_{\alpha_{antiMI}, x, t}(\alpha_2)}$$

$$c^*_{\alpha_2}(\alpha_{antiMI}, x, t) = \frac{attr^*_{\alpha_{antiMI}, x, t}(\alpha_2)}{attr^*_{\alpha_{antiMI}, x, t}(\alpha_1) + attr^*_{\alpha_{antiMI}, x, t}(\alpha_2)}$$