

teorema

Vol. XXXVII/2, 2018, pp. 31-54

ISSN 0210-1602

[BIBLID 0210-1602 (2018) 37:2; pp. 31-54

Knowledge First and Rational Action

Christian J. Feldbacher-Escamilla

RESUMEN

La epistemología del *conocimiento primero* trata de resolver los callejones sin salida a los que ha llegado la investigación epistémica como resultado de haber puesto la *creencia primero*. Este nuevo paradigma tiene en su centro, una metodología abductiva y una tesis sobre la prioridad explicativa del *conocer*. De aquí se derivan varias normas del conocimiento como lo son, por ejemplo, la norma de la acción del conocimiento, que enuncia que uno debe actuar solamente en base a lo que uno sabe. Mientras esta norma está en lo correcto en cuanto a los estándares y la culpabilidad epistémica, se opone a la justificación tradicional de las leyes de la creencia y las acciones racionales. Este artículo resume histórica y sistemáticamente los paradigmas de la *creencia primero*, así como del *conocimiento primero*. Además, se explican las afirmaciones centrales en relación con la acción y se presenta y se justifica una formulación exacta de la norma de la acción del conocimiento. Esta norma evita los problemas mencionados antes, mientras que simultáneamente mantiene las ventajas de poner por delante altos estándares epistémicos.

PALABRAS CLAVE: *epistemología del conocimiento primero, normas del conocimiento, teoría de la decisión, probabilismo, normatividad epistémica, culpabilidad epistémica.*

ABSTRACT

Knowledge first epistemology aims at dissolving deadlocks in epistemic research that seem to be caused by putting *belief first*. This new paradigm has an abductive methodology and a thesis on the explanatory priority of *knowing* in its core. From this it derives several knowledge norms as, e.g., the knowledge norm of action, which states that one should only act on what is known. While this norm gets things right with respect to epistemic standards and epistemic blamability, it runs counter to the traditional justification of the laws of belief and rationality of actions. This paper gives a historical and systematical overview of the *belief first* and *knowledge first* paradigms. The central claims regarding action are explicated and an exact formulation of the knowledge norm of action is provided and justified. This norm avoids the mentioned problems while at the same time maintaining the advantages of putting forward high epistemic standards.

KEYWORDS: *Knowledge First Epistemology, Knowledge Norms, Decision Theory, Probabilism, Epistemic Normativity, Epistemic Blamability.*

I. INTRODUCTION

The main idea of *knowledge first* epistemology is to reverse the traditional direction of explanatory priority. Whereas in traditional epistemology very often normative phenomena regarding knowledge were explained by normative considerations on belief, *knowledge first* epistemology tries to explain phenomena regarding belief on the basis of normative considerations on knowledge. The first highly influential elaboration of this project was performed in *Knowledge and its Limits*: “If I had to summarize this book in two words, they would be: knowledge first” [Williamson (2000), preface, p. v]. Since this shift from belief towards knowledge is justified by help of fruitfulness considerations, the background methodology of this approach directs towards a choice of the best explanation of epistemic desiderata, which is simply abductive reasoning applied in philosophy.

Knowledge first epistemology is best characterised as a new paradigm of epistemology or as a new epistemological research programme. It consists of the mentioned abductive methodology and a core thesis on the primacy of knowledge. From this thesis several knowledge norms concerning other epistemic notions are derived: A knowledge norm of belief, evidence, justification, assertion, and action or decision. Since knowledge is considered to be the epistemic standard we should start with, all the other epistemic attitudes are supposed to approximate this standard [cf. Benton (2014)]. For this reason, the underlying knowledge norm is as follows: One ought to φp only if one knows p . I.e.: One ought to belief, take as evidence, justify by help of, assert that, and act on p only if one knows p .

These knowledge norms are quite controversially discussed in the literature [cf. McGlynn (2014)]. They set very high epistemic standards for which reason it is questioned whether one can satisfy them. So, e.g., regarding action it seems to be problematic to assume that one can rationally act only on something one knows. In this paper we aim at a detailed explication of such norms of action and try to compare them regarding their advantages and disadvantages. Since the *belief first* rival emerged from a critique of a precursor of the *knowledge first* paradigm, namely logical probabilism, we also want to locate the new approach in the tradition of thoughts by providing a historical overview and explicating the dialectic between these paradigms.

The structure of the paper is as follows: In section II we describe the approach to rational action of the *belief first* paradigm and its main advantages. In section III we give a sketch of an early *knowledge first* epistemology, namely logical probabilism, and its approach to rational

action. We highlight some problems commonly put forward, but mention also some advantages of this approach. In section IV we outline the modern *knowledge first* approach to action of [Williamson (2000), (2017)]. There we also state the main advantage of this approach, namely being capable of explaining epistemic blamability, i.e. the possibility to be blamed, although one's action might be regarded as rational. In section V we discuss some objections to this approach, explicate it further and show how the explicated *knowledge first* approach to action can avoid the objections. We conclude in section VI.

II. RATIONAL ACTION AND *BELIEF FIRST*

In decision and utility theory the standard way of thinking is that, at least from a normative standpoint, the action of an agent is completely determined by her beliefs and desires, values or utilities. If we assume that F is the set of alternative actions f_1, \dots, f_n on the set S of possible states of the world available on a given occasion, then *expected utility theory* proposes to decide for that alternative action which has greatest expected utility [cf., e.g., Schurz (2013), pp. 167f]. If there is more than one such alternative available, then one might randomly pick among them, but for simplicity of expression we assume that there is always exactly one such action. To calculate the expected utility of an action means to just sum over the products of the utilities of the action on a state $s \in S$ of the world and the probability of that state s :

$$eu(f_i) = \sum_{s \in S} u(f_i(s)) \cdot Pr(s) \quad (eu)$$

So, *eu* combines the utility of the consequences of an action (outcomes $f_i(s)$) with their probabilities. A simple decision rule of expected utility theory can be stated then as follows:

$$decision(F) = f_i \text{ iff } eu(f_i) = \max_{f_j \in F} (eu(f_j)) \quad (decision)$$

Now, as we can see, an action (*decision*) is determined by expected utilities (*eu*) and expected utilities in turn are determined by utilities (u) and probability estimations or degrees of belief (Pr). So, this approach bases actions

and decisions in beliefs, i.e. it is an approach to action and decision within the *belief first* paradigm. Sloppily speaking it provides a (*desire and*) *belief norm of action*: You ought to act in accordance with your credences (and desires).

Orthodox decision theory does not stop here. It goes on with an explanatory reduction of *Pr* via linking degrees of belief to, e.g., betting behaviour. So, e.g., Frank P. Ramsey claimed that:

As soon as we regard belief quantitatively [...] we seek to know what is the difference between believing more firmly and believing less firmly [...] The difference seems to me to lie in how far we should act on these beliefs [...] The kind of measurement of belief with which probability is concerned [...] is a measurement of belief *qua* basis of action.” [cf. Ramsey (1926/1950), pp.170f]

A simple way of grounding *Pr* behaviouristically in *decision* among this line is to identify *Pr* with the highest/lowest odds one accepts as fair such that one is willing to bet at them on/against p . Odds are characterised as ratios $n : m$ (n, m positive), where n might represent one’s estimation of favourable cases, and m that of unfavourable cases. Given these assumptions, the basic set of actions B consists of $b_{1,n,m}$: betting on p at odds $n : m$, $b_{2,n,m}$: betting against p at odds $n : m$, and $b_{3,n,m}$: abstaining from betting at odds $n : m$. If we consider just the possible states of the world $S = \{p, \neg p\}$, then we can identify the utilities of betting on or against or abstaining with the following payoffs:

$B :$	$S :$	p	$\neg p$
$b_{1,n,m}$	bet on p	$u(b_1(p)) = +m$	$u(b_1(\neg p)) = -n$
$b_{2,n,m}$	bet against p	$u(b_2(p)) = -m$	$u(b_2(\neg p)) = +n$
$b_{3,n,m}$	abstain from betting	$u(b_3(p)) = 0$	$u(b_3(\neg p)) = 0$

Given these payoffs it seems reasonable to apply the following principle of determining degrees of belief:

$$Pr(p) = r \text{ iff}$$

there are $n, m (\geq 0)$ such that $r = \frac{n}{n+m}$ and

$$\text{decision}(B) = b_{1,n,m} \text{ or } \text{decision}(B) = b_{2,m,n} \tag{Pr}$$

and for any $n^* : m^* > n : m$ $\text{decision}(B) \neq b_{1,n^*,m^*}$
and for any $n^* : m^* < n : m$ $\text{decision}(B) \neq b_{2,n^*,m^*}$

E.g., if one believes that a particular ordinary die is unbiased towards three, i.e. one's degree of belief in the occurrence of a three on the roll of a die is $1/6$, then one ought to accept a bet up to the odds $1x : 5x$ (for any positive x) on the occurrence of a three. So, if a bookie offers a bet, e.g., with odds $1\text{€} : \geq 5\text{€}$, one is rational in adding to the pot 1€ and getting out $\geq 6\text{€}$ in case three shows up while getting out nothing in case three does not show up. On the other hand, one is also rational to accept a bookie's bet on the occurrence of a three (so oneself bets against the occurrence of a three) starting with odds of $1x : 5x$ (for any positive x), e.g., all bets with odds $\geq 1\text{€} : 5\text{€}$. So, one is rational in adding to the pot 5€ and getting out $\geq 6\text{€}$ in case three does not show up while getting out nothing if three shows up.

Just to illustrate this analysis of *rational action* in the *belief first* approach further, we also want to mention the real world example of playing lottery: It is easy to see that according to this interpretation joining an ordinary lottery game is not rational for most of the participants buying a lot (whereas the actions of the bookies are clearly rational): Most of the participants buying a lot would agree that in a lottery where one chooses, e.g., 6 numbers from 49, the chances and by this also their degrees of belief to win are $\approx 1/14.000.000$. However, in fact the average payout for a winner in such a lottery is a little bit less than 50.000€ , so these lottery participants are *de facto* accepting bets at odds $\approx 1\text{€} : 50.000\text{€}$: $b_{1,\approx 1:50.000}$, although according to their degrees of belief they should accept only bets up to odds $\approx 1\text{€} : 14.000.000\text{€}$ and hence, according to (*Pr*) the rational action would be to either turn to a bookie oneself: $b_{2,\approx 1:50.000}$ or abstain from betting: $b_{3,\approx 1:50.000}$.

Now, grounding *Pr* in *decision* might look circular at first glance and, given the decision rule above (*decision*), it clearly is. However, it is not so in a vicious way: The idea is to ground *decision*(*F*) in *Pr* amongst others, and *Pr* in turn is grounded in *decision*, but not in *decision*(*F*), rather in *decision*(*B*) where *B* represents another set of actions, namely actions of accepting or declining bets. In such a way betting might be seen as a means of "measuring beliefs" [cf. Ramsey (1926)/(1950), p. 176]. Note that by this the *belief norm of action* turns to an *action norm of belief*: You ought to belief in accordance with your decisions and actions about bets, lotteries etc. To sum up, this is how an orthodox *belief first* approach analyses *rational action*:

The Belief First Approach to Rational Action

An action $f \in F$ is *rational* iff it is in accordance with (*decision*), where the expected utilities eu in (*decision*) are in accordance with (eu), and where the degrees of belief Pr in (eu) have to be in accordance with (Pr) on the corresponding betting action space B .

$$\text{Schema: } decision(B) \Rightarrow Pr \Rightarrow eu \Rightarrow decision(F)$$

$$\begin{array}{c} \uparrow \\ u \end{array}$$

Advantages of this approach are manifold. Here we want to mention the perhaps two most famous ones: There are, e.g., the famous representation theorems of expected utility theory which allow one to link quite naturally constraints on actions given ordinal (comparative) information with constraints on actions given cardinal (quantitative) information. So, e.g., if \leq_p is a relation on the set of actions F which satisfies a set of plausible axioms for preferences, then one can prove that for all $f_i, f_j \in F$ it holds: $f_i \leq_p f_j$ determines a unique Pr and u such that $eu(f_i) \leq eu(f_j)$ and the other way round [infamous are, e.g., von Neumann and Morgenstern (1944)/(1953); and Savage (1972)]. Such representation theorems suggest to consider expected utilities as a quite straight-forward expansion of ordinary preference relations.

That expected utilities eu are a straightforward expansion of preferences is a quite important insight since, e.g., many impossibility results of social choice theory on the aggregation of preference relations disappear on a quantitative scale [cf., e.g. Gaertner (2009), ch. 7]. So, e.g., Kenneth Joseph Arrow's infamous result of (1951)/(1963) about the impossibility of a preference aggregation rule that satisfies constraints of being universally applicable, unanimity preserving, informational parsimonious, and non-dictatorial disappears on the quantitative scale [cf., e.g., the characterisation result of linear aggregation in Genest and Zidek (1986)]. So, in general grounding action on belief which is in turn grounded on action again provides a justified incentive, e.g., for avoiding social choice impossibilities by switching to a quantitative level.

The second big advantage of this *action norm of belief which in turn provides a normative standard of action* that needs to be mentioned here is its capability of justifying the standard laws of graded belief. So, e.g., the infamous *Dutch book* arguments show that putting forward constraints on actions regarding betting and accepting lotteries combined with an *action norm of belief*

as outlined in (Pr) allows for a justified application of the Kolmogorov-axioms to degrees of belief:

$$(Pr1) \ Pr(p) \geq 0$$

$$(Pr2) \ Pr(\top) = 1$$

$$(Pr3) \ Pr(p \vee q) = Pr(p) + Pr(q) \text{ given } p \vdash \neg q$$

In a nutshell (Pr) allows for a simplified Dutch book argument: Non-negativity $(Pr1)$ follows immediately from (Pr) since the ratio of positive numbers is positive. Normalisation $(Pr2)$ follows from (Pr) and the assumption that one should not be prone to Dutch books, i.e. a set of bets whose payoff *in toto* is negative regardless of the outcome: $Pr(\top) > 1$ is excluded by the ratio-formula; If $Pr(\top) = 1 - r$ ($r > 0$), then you might be offered to bet against \top at odds $1: r/(1 - r)$: $b_{2,1:r/(1-r)}$; since the outcome is always \top , you will never get out anything, but have to pay 1. Regarding finite additivity $(Pr3)$ just consider the following possibility: If $Pr(p \vee q) < Pr(p) + Pr(q)$, although $p \vdash \neg q$, then a bookie can offer the following bets for the agent: $b_{1,Pr(p):1}$ on p , $b_{1,Pr(q):1}$ on q , $b_{2,1:Pr(p \vee q)}$ against $p \vee q$. Since $p \vdash \neg q$, this allows for the following possibilities with the following payoffs for the agent:

S		Payoffs			Net Payoff
$\neg p$	$\neg q$	$-Pr(p)$	$-Pr(q)$	$+Pr(p \vee q)$	< 0
p	$\neg q$	$1 - Pr(p)$	$-Pr(q)$	$1 - Pr(p \vee q)$	< 0
$\neg p$	q	$-Pr(p)$	$1 - Pr(q)$	$1 - Pr(p \vee q)$	< 0

Since by assumption $Pr(p \vee q) < Pr(p) + Pr(q)$ for all possible states the net payoff is negative, which means that by such a set of bets the agent is Dutch booked. By reversing the betting direction in the above bets ($b_1 \rightarrow b_2$ and $b_2 \rightarrow b_1$) one can also rule out $Pr(p \vee q) > Pr(p) + Pr(q)$ as rational. So, if one interprets degrees of belief behaviouristically as specific betting behaviour in accordance with (Pr) and accepts as a rationality constraint to avoid the possibility of being Dutch booked, then $(Pr1)$ – $(Pr3)$ have to hold. The so-called *converse Dutch book theorem* shows that $(Pr1)$ – $(Pr3)$ are not only necessary for not being Dutch booked, but also sufficient [cf. Hájek (2005), p. 141].

To sum up: *Belief first* epistemology allows for explaining the rationale behind an action via expected utilities. Representation theorems show that

such an explanation equals a straightforward expansion from the comparative realm of preferences to the quantitative realm of utilities. Furthermore, *belief first* can be even grounded in basic norms of actions, namely intuitions on adequate betting behaviour. In the next section we will present a precursor of the *knowledge first* account to rational action: logical probabilism. As we will see, the quantitative version of *belief first* epistemology described here arose in the course of a critique of this approach.

III. RATIONAL ACTION AND A PRECURSOR OF *KNOWLEDGE FIRST*: LOGICAL PROBABILISM

As we have indicated above, the traditional epistemic account of *belief first* was formally spelled out by authors like Bruno de Finetti (1972), Richard C. Jeffrey (1983), Ramsey (1926)/(1950), and Leonard J. Savage (1972). Also regarding knowledge these authors put belief first. So, e.g., Ramsey claimed:

I have always said that a belief was knowledge if it was (i) true, (ii) certain, (iii) obtained by a reliable process [...where certainty consists in a high enough degree of belief and] a belief obtained by a reliable process must be caused by what are not beliefs in a way or with accompaniments that can be more or less relied on to give true beliefs, and if in this train of causation occur other intermediary beliefs these must all be true ones" [cf. Ramsey (1929)/(1950), p. 258].

What we find here is one of the first exact *justified (iii) true (i) belief (ii)* account of knowledge in quantitative terms, where a proposition is not just believed simpliciter, but believed to a high enough degree. It is interesting to note that especially Ramsey directed his approach against an approach that might be considered to be a forerunner of modern *knowledge first* epistemology, namely John Maynard Keynes' theory of probabilities which more or less explicitly puts forward a *knowledge norm of belief*:

Part of our knowledge we obtain direct [by direct acquaintance, i.e. experience and perception]; and part by argument. The *Theory of Probability* is concerned with that part which we obtain by argument, and it treats of the different degrees in which the results so obtained are conclusive.

The terms *certain* and *probable* describe the various degrees of rational belief about a proposition which different amounts of knowledge authorise us to entertain. [...] Given the body of direct knowledge which constitutes

our ultimate premisses, this theory tells us what further rational beliefs, certain or probable, can be derived by valid argument from our direct knowledge. This involves purely logical relations between the propositions which embody our direct knowledge and the propositions about which we seek indirect knowledge. [cf. Keynes (1921), ch. 1, sec. 1f].

Keynes, one of the first proponents of a theory of *logical probability*, suggested to put that what is known first, and then explain by logical relations between propositions what is probable, certain, or what can be believed with a specific degree of belief. Perhaps most prominently in a similar line is Rudolf Carnap's seminal (1962) which initiated a research programme upheld until recently by, e.g., Patrick Maher (2001). The logical relations these theories of logical probability mainly want to cash out derive from a so-called *principle of indifference*, introduced into the mathematical Theory under the label *The Principle of Non-Sufficient Reason* [cf. Keynes (1921), p. 44]. Different versions of such a principle have been employed. The core idea underlying this principle can be *naïvely* stated as follows:

The Principle of Indifference asserts that if there is no known reason for predicating of our subject one rather than another of several alternatives, then relatively to such knowledge the assertions of each of these alternatives have an equal probability [Keynes (1921), p.45].

Naïve [cf. van Fraassen (1989), p. 303] is such a statement of this principle in the same way as an unrestricted statement of comprehension is in *naïve set theory*: Without further restrictions it leads to inconsistency. Regarding the principle of indifference very influential was Joseph Bertrand (1888)/(1907) who discussed problems of this principle today known as *Bertrand's paradox*. He highlighted that applying such a principle to different descriptions of one and the same event results in different probability statements of the event. This problem can be easily illustrated by the following example of van Fraassen (1989), p. 303: Assume a precision tool factory produces iron cubes c_i . Now, assume further that we have three different languages/theories that allow us to express another fact we know about the c_i s: In language/theory L_1 we can express that the edge length is within a specific interval, let us say $\forall i : a(c_i) \in [0, 2]cm$. In other terms of another language L_2 we can express this fact equivalently as the claim that the area of a side is within a specific interval: $\forall i : a(c_i) \in [0, 4]cm^2$. And yet in other terms of yet another language L_3 we can equivalently claim that the volume is within a specific interval: $\forall i : v(c_i) \in [0, 8]cm^3$. Now, given this knowledge as starting point, a *naïve* application of the principle of indifference leads to the following “degrees of rational

belief about a proposition which different amounts of knowledge authorise us to entertain”:

$$Pr(l(c_j) \in [0, 1]cm \mid \forall i: l(c_i) \in [0, 2]cm) = 1/2,$$

$$Pr(a(c_j) \in [0, 1]cm^2 \mid \forall i: a(c_i) \in [0, 4]cm^2) = 1/4, \text{ and}$$

$$Pr(v(c_j) \in [0, 1]cm^3 \mid \forall i: v(c_i) \in [0, 4]cm^3) = 1/8.$$

Note that $l(c_j) \in [0, 1]cm$, $a(c_j) \in [0, 1]cm^2$, and $v(c_j) \in [0, 1]cm^3$ describe the same event/property of cube c_j equivalently. However, a *naïve* application of the principle of indifference licenses contradicting probability statements.

Keynes and adherents of the programme of logical probability faced problems as the one presented above and were in need of spelling out restrictions of the *naïve principle of indifference* that allow for avoiding inconsistency. Shortly, we will speak about a further problem of this account. However, since we are interested on the relation between knowledge and action here, we want to address the question about this relation first: What is the logical probabilist’s approach to action in terms of knowledge? Regarding this question logical probabilists seem to be pretty much in line with the classical *belief first* approach (cf., e.g., the hints on applying probability theory within philosophy and ethics of Keynes (1921), p. 353): An action is rational if it maximises expected utilities. However, maximising the expectation value is not based on belief which is furthermore based on betting behaviour. Rather, it is based on belief that is assumed to be “authorised” logically by what we know, our evidence. So, the logical probabilist’s *knowledge norm of action* might be formulated as follows:

The Logical Probabilities Approach to Rational Action

An action $f \in F$ is *rational* iff it is in accordance with (*decision*), where the expected utilities eu in (*decision*) are in accordance with (eu), and where the Pr in (eu) are the “degrees of rational belief about a proposition which different amounts of *knowledge* authorise us to entertain” [for the last part cf. Keynes (1921), chpt.1, sect.1f].

$$\text{Schema: } \textit{knowledge} \Rightarrow Pr \Rightarrow \begin{array}{c} eu \Rightarrow \textit{decision}(F) \\ \uparrow \\ u \end{array}$$

So far, a sketch of the programme of logical probability putting knowledge first and degrees of belief and action second. An advantage of this approach consists in the high epistemic standard it promises to provide since reasoning and rational action starts with knowledge and then proceeds by logical principles in a wide sense. However, there is a fundamental problem of this approach which is stressed especially by adherents of the *belief first* camp. Again, Ramsey's critique of Keynes' theory describes the problem very well and marks a pattern that we find also in the more modern debate, as we will show below (section V).

The critique concentrates on the fact that putting *knowledge first* and then deriving rationality constraints for degrees of belief given the known propositions lacks justification. As Ramsey puts it:

I do not see what these inconclusive logical relations can be or how they can justify partial beliefs. [. . .] The logical relation which justifies the [deductive] inference is that the sense or import of the conclusion is contained in that of the premisses. But in the case of an inductive argument this does not happen in the least; it is impossible to represent it as resembling a deductive argument and merely weaker in degree [Ramsey (1929)/(1950), pp. 185f].

Ramsey then goes on to highlight that his *belief first* approach does not suffer from this problem: Axioms of Keynes' theory that lack a logical justification like the *principle of indifference* are no longer necessary in a theory of partial belief or degrees of belief that bases such degrees on actions. While the former puts forward constraints on such axioms, the latter does not put forward any such constraints whatsoever:

[Basing degrees of belief on actions] is simply bringing probability into line with ordinary formal logic, which does not criticize premisses but merely declares that certain conclusions are the only ones consistent with them. To be able to turn the Principle of Indifference out of formal logic is a great advantage; for it is fairly clearly impossible to lay down purely logical conditions for its validity, as is attempted by Mr Keynes [Ramsey (1929)/(1950), p. 189].

The above-mentioned problem of spelling out a coherent version of the principle of indifference and how to "logically" justify such principles is generally regarded as one of the main problems of the theory of logical probabilities. Due to its increased number of parameters and its lack of generality it is commonly considered to be a degenerative research programme [cf., e.g., Spohn (1981), p. 50]. However, as we have shown

here, it can be considered as an approach within the *knowledge first* paradigm and it served as a stimulus of the predominant *belief first* paradigm as outlined in section II. In the next section we present *the* modern programme of putting *knowledge first*.

IV. RATIONAL ACTION AND KNOWLEDGE FIRST

As we have shown in the preceding section, also in the formal and quantitative realm *knowledge first* epistemology has already a quite long tradition. However, the *knowledge first* paradigm underwent a revival and started again from a completely different angle with Timothy Williamson's seminal (2000). As Carter, Gordon, and Jarvis (2017a) put it: "Timothy Williamson is the founder of the knowledge-first movement and has been its principal flagbearer."

In Williamson (2000), many epistemic implications of putting knowledge first are discussed. The main motivation for this approach is based on the fact that Edmund L. Gettier's paper on the traditional conditions of knowledge (1963) caused a whole industry of approaches that try to save the so-called *justified true belief (JTB)* account of knowledge [for a short discussion cf., e.g., Boongaling (2017)]: Gettier prominently argued against an analysis of *knowledge* in terms of *JTB* by help of examples where an agent has justified true belief, but where one still refrains from attributing knowledge to the agent. For this reason, the *JTB* analysis was expanded to an *JTB + X* analysis of knowledge, where *X* stands for further conditions like, e.g., an anti-luck condition which again might be spelled out via other notions as, e.g., the notion of *sensitive belief* or that of *safe belief* [cf. Pritchard (2014)]. However, in order to avoid further counterexamples, the conditions *X* become more and more complicated which is for some authors an indicator of a degeneration of the *JTB + X* research programme. So, it was argued, e.g., that for any further condition *X* a Gettier-style counterargument to an *JTB + X* analysis of knowledge can be constructed which suggests that any further *X* is either redundant, because it is already implicitly contained in *JTB*, or circular inasmuch as it presupposes the notion of *knowledge* already [cf. Zagzebski (1994)]. This led modern *knowledge first* epistemologists as, e.g., Williamson to the educated and methodologically backed up guess that *knowledge* might not be fruitfully analysed, but better taken as starting point of all epistemic endeavors.

In order to avoid this kind of Gettier-deadlock [cf. Williamson (2016), pp. 275f] *knowledge first* epistemology does not provide an analysis of *knowledge* in terms of necessary and jointly sufficient conditions, but rather uses it for making unificatory claims. Perhaps one of the most influential claims within this approach is the so-called *knowledge first thesis*:

Knowing is the most general truth-entailing mental attitude, the one you have to a proposition if and only if you have any truth-entailing mental attitude to it at all [cf. Williamson (2011), pp. 215f].

Mental attitudes are attitudes as, e.g., desiring, wanting, hoping, seeing, fearing, thinking, believing, being aware, remembering, knowing, etc. Clearly non-factive/not truth-entailing are, e.g., attitudes like desiring, wanting, hoping, believing: Having the attitude towards something/a proposition does not imply that something/the proposition is the case/true. Clearly factive/truth-entailing, at least in considering strong versions of these attitudes (so-called *fact attitudes* in contrast to *object attitudes* [cf. Williamson (2011), p. 216]), are, e.g., seeing, remembering, knowing: To see *that* a leaf is green implies the truth of ‘The leaf is green.’ Similarly for *remembering that* and *knowing that*.

The knowledge first thesis can be itemised as follows [cf. Williamson (2000), preface, p. 39]:

(KFT1) If φ is a [factive mental state operator], from ‘ $S \varphi$ that \mathcal{A} ’ one may infer ‘ \mathcal{A} ’.

(KFT2) ‘Know’ is a [factive mental state operator].

(KFT3) If φ is a [factive mental state operator], from ‘ $S \varphi$ that \mathcal{A} ’ one may infer ‘ S knows that \mathcal{A} ’.

The thesis might be also described via the following schema:

$$\overbrace{\Box(K(p) \rightarrow p)}^{(KFT2)} \quad \& \quad \overbrace{\Box(M(p) \rightarrow p) \rightarrow \Box(M(p) \rightarrow K(p))}^{(KFT1) \text{ and } (KFT3)}$$

So, if M is a modal operator representing a factive mental attitude towards proposition p ($\Box(M(p) \rightarrow p)$), then having this attitude towards p , i.e. $M(p)$, implies also knowing p , represented by the modal

operator K : $K(p)$. Here \Box might be interpreted as conceptual necessity (satisfying factivity itself, i.e. the modality of system T of modal logics).

Now, one might argue for the unificatory power of the knowledge first thesis as follows: Our epistemic goal is the truth, let us say proposition p . Now, having a factive mental attitude towards p (i.e. $\Box(M(p) \rightarrow p)$ and $M(p)$) is not only a sufficient means to achieve the goal, but regarding our epistemic capabilities the only means to achieve the goal. Hence, we *ought* to have a factive mental attitude ($\Theta M(p)$, where Θ expresses epistemic normativity) towards the truth (p). Now, since M is factive, according to the *knowledge first thesis* having $M(p)$ conceptually implies knowing the truth: $K(p)$ ($\Box(M(p) \rightarrow K(p))$). Hence, we also ought to know the truth: $\Theta K(p)$. More generally: Whenever we aim at the epistemic goal truth, we also aim at knowledge. Hence we can infer that *knowing is the most general epistemically obligatory mental attitude, the one you have to a proposition if and only if you have any epistemically obligatory mental attitude to it at all*.

According to the knowledge first thesis knowing is the most general factive attitude. This suggests to consider it as a *species* of all other factive mental attitudes: *remembering that is knowing through memory* and *seeing that is knowing through vision* etc. But what about non-factive attitudes towards p ? E.g., what about believing that p or deciding on basis of p ? Since knowing is the most general factive mental attitude one can have and one always aims at if one aims at the truth, one can formulate a set of *knowledge norms*. E.g.: The knowledge norm of *evidence* states that one ought to take as evidence exactly what one knows (this is the famous $E=K$ -thesis/norm). Regarding belief (B), *knowledge first* epistemologists state a *knowledge norm of belief*: “believe that p only if you know that p ” [cf. Williamson (2011), p. 214]. Schematically:

$$\Theta(B(p) \rightarrow K(p))$$

Clearly, such a norm puts forward very high epistemic standards for belief. But this does not imply that there is no way of assessing beliefs that are defective in the sense that they fall short of being knowledge. So, e.g., comparatively speaking if, e.g., $B(p \& q)$ and $B'(p \& q)$, where $K(p \& q)$ while $K'(p) \& \neg K'(q)$, then clearly B approximates K better than B' approximates K' and hence B satisfies the knowledge norm of belief better than B' does.

The same line of argumentation can be also applied for the attitude of deciding on basis of p . Clearly, such an attitude is not factive:

Genuine mental states play a role in causal explanations of rational action. Such explanations cite the agent's beliefs and desires. They do not cite the agent's knowledge as such, for even if some of those beliefs do in fact constitute knowledge, that is irrelevant to their role in causing the action. If you want a drink of water and believe that your glass contains water, then other things being equal you will drink from your glass, whether or not you really know that it contains water [cf. Williamson (2011), p.213]

So, *de facto* deciding on basis of p might be defective, namely in case p is false. This is due to the fact that decisions *are* based on beliefs. However:

from a normative perspective, knowing is arguably more relevant to rational action than believing is. Suppose that, in deciding whether to take certain pills, I treat as a reason for taking them that they will alleviate my hay fever symptoms. If you point out that I don't know whether they will alleviate my symptoms, you reveal a defect in my decision-making with respect to that premise. [...] If I don't know the premise but have strong evidence that I do know it, that may be a good excuse for having treated it as a reason, but that does not make the decision-making non-defective, otherwise no excuse would be needed. [...] Urgency is another excuse for defective reasoning; it does not remove the defect [cf. Williamson (2011), p. 214]

Also, an argument from analogy can be put forward in favour of this normative stance: There seems to be a fruitful analogy between knowing that p and acting on p : It seems to be plausible to map these notions as follows [cf. Williamson (2017), sect.1]:

knowledge	is analogous to	action
belief	is analogous to	intention
truth	is analogous to	success
falsity	is analogous to	failure

Regularities between these notions are as follows [cf. Williamson (2017), sect. 3f]: Knowledge entails true belief. But not the converse. Action entails successful intention. But not the converse. Furthermore, it does not make much sense to try to understand intention without reference to action. Analogously, if knowledge is to belief as action is to intention, it does not make much sense to try to understand belief without reference to knowledge — that is a core claim of *knowledge*

first epistemology. Given these similarities, it seems reasonable to formulate a *knowledge norm of action* analogously to the *knowledge norm of belief* presented above [cf. Benton (2014), sect.2; Hawthorne and Stanley (2008), p. 577; Lackey (2010); McGlynn (2014), p. 132]:

$$\mathcal{O}(D(p) \rightarrow K(p))$$

where ‘ $D(p)$ ’ stands for ‘deciding to act on p ’, ‘treating a proposition p as a reason for action’.

Roughly speaking, this is the current *knowledge first* approach to rational action. We will explicate this approach further in the next section. But before that we want to highlight two advantages of it: First, as was also the case in the logical probability programme, the modern *knowledge first* approach sets high epistemic standards since it demands as starting point of reasoning and action or decision knowledge. However, contrary to the logical probability approach it softens the rules that can operate on this starting point: The aim is not to apply “logical” rules only, but any rules that allow for approximation of knowledge. Due to this the modern research programme seems to be not prone of degenerative fitting and fine-tuning from the start on.

Second, the *knowledge first* approach allows for an explanation of arguments from *blamability* and *excusability*. To see this, consider the following example: Hannah, the wife of Mordechai, remarries after waiting five years for the return of her husband, who has been missing in battle. Now, it seems that Mordechai “might legitimately complain to Hannah that she shouldn’t have remarried without knowing that he had died. It is reasonable for Mordechai not to be satisfied with the excuse that Hannah had a justified belief that he was dead” [cf. Hawthorne and Stanley (2008), pp. 573f — attributed to Jim Pryor]. The above knowledge norm of action allows for identifying this defect in Hannah’s decision. She deviated from the epistemic norm by basing her decision on justified belief that is not knowledge. So, justified belief alone does not provide a reason and does not allow for an (absolute) excuse. Note that the quotation of [Williamson (2011)] above suggests a little bit different wording, while still expressing the same fact: Justified belief might provide an excuse, but the decision still can be called ‘defective’.

Putting knowledge first can explain why one can be blamed, although one’s action was rational. Things are different for the *belief first* approach: If one starts with belief, then there seems to be little space for such an explanation: If we assume that Hannah maximised the expected utilities, and if max-

imisation of expected utilities is all one has to strive for, then there is nothing one needs an excuse for or one can be blamed for.

In the following section (V) we will discuss some objections to this approach, explicate it further, and show how the explicated knowledge norm of action can deal with the objections.

V. EXPLICATION, OBJECTIONS AND REPLIES

As we have seen above, the *knowledge norm of action* states that one ought to act or decide on p only if one knows p . There are several problems put forward against this *knowledge norm of action*. We will concentrate on the two most common problems [cf. Kaplan (2009)]. Doing so allows us also to simultaneously explicate the norm further. First, there is the problem that such a norm seems to be unable to explain most of our actions and decisions since most of them are based on belief that falls short of being knowledge. And second, there is the problem that such a rule leaves the question of how to justify laws of belief (degrees of belief) open.

Let us begin with the first problem: The knowledge norm of action states that one ought to treat a proposition p as a reason for action or decision only if one knows p ($\mathcal{O}(D(p) \rightarrow K(p))$). Now, on most occasions we lack knowledge about the exact state of the world, although we might have degrees of belief about the possible states of the world. However, according to the knowledge norm of action we cannot employ this information, since it falls short of being known. But this seems counterintuitive. Consider the following example:

Suppose that you are suffering discomfort, and so go to the doctor. After the requisite examination and tests are completed, she tells you that your discomfort is not a sign of any dangerous condition. For all that, she continues, there is a procedure you can undergo that, if you have condition C, will completely alleviate the discomfort. If, however, you do not have condition C, your discomfort will remain as it is. The procedure is, however, quite expensive. The choice as to whether to undergo the procedure, she tells you, is up to you [Kaplan (2009), p. 135].

It seems clear and rational to ask the doctor for the probability of having condition C before one makes a decision. Although this is full in agreement with a *belief first* approach to action and decision, it seems to be not justified within a *knowledge first* approach since the doctor does not *know* whether one has condition C or not [cf. Kaplan (2009), p. 136].

However, we think that this is not a very benevolent interpretation of the *knowledge first* approach: Similarly, as the knowledge norm of belief ($\mathcal{C}(B(p) \rightarrow K(p))$) justifies the norm that belief ought to approximate knowledge, one can also put forward the norm that our actions and decisions based on belief ought to approximate our actions and decisions based on knowledge. The former actions and decisions are “defective” inasmuch as they license blaming, whereas the latter are not defective and do not allow for blaming. So, in general *knowledge first* epistemology allows for decision in accordance with the norm of *belief first* epistemology as described in (*decision*). But contrary to *belief first* which rationalises an action just in comparison with all available alternatives, *knowledge first* puts forward a very high standard for comparison, namely knowledge.

Let us make this clear by help of a toy model: Assume, as in section II, F to be the set of alternative actions f_1, \dots, f_n on the set S of possible states of the world. Then one element of S is the actual state, let us label it with ‘ p ’. Now, according to the knowledge norm of action every consideration of the probability and utility of an $s \neq p \in S$ in an action or decision falls short of using knowledge and by this is defective. So, a non-defective decision takes into account only the utilities of an action f_i on p . Following (*en*) this is technically equivalent to having degrees of belief $Pr(p) = 1$ and $Pr'(s) = 0$ for all $s \neq p \in S$. Deviating from such an extreme distribution might still approximate a knowledge-based decision, but is still defective. So, given the definitions (*en*) and (*decision*) we can define a measure for defectiveness of a decision by simply taking the distance of the agent’s degree of belief in p from 1:

$$\begin{aligned} \text{def}(\text{decision}) &= 1 - Pr(p) \\ &\text{(where } p \text{ is the actual world state)} \end{aligned} \quad (\text{def})$$

Clearly, not all beliefs that take the actual state (p) as certain make up for knowledge. So, there are decisions that are defective although they base their decision in certainty of p and have a defectiveness measure of 0. Hence, $\text{def}(\text{decision}) = 0$ does not imply that *decision* is non-defective at all. It just states that *decision* is not defective with respect to the underlying degree of belief. To put it differently: $\text{def}(\text{decision}) = 0$ is only a necessary condition for *decision* to be non-defective, however, it is not a sufficient one.

Although more needs to be said and also can be said on this issue, there will be an element of an exact characterisation of the defectiveness of a decision that is irreducible in principle: To spell out the exact conditions of a *decision*’s failing to be non-defective due to its being based on belief (B)

bilism can boast, leaves me feeling secure in my probabilism. What I do not see [. . .] is how Williamson can feel secure in his [Kaplan (2009), p. 139].

Mark Kaplan takes this as a reason for speaking of ‘Williamson’s *casual* approach to probabilism’. Note that this complaint is reminiscent of Ramsey’s critique of Keynes’ *knowledge first* stance within the programme of logical probabilities (cf. section III). However, an important difference is that, whereas the programme of logical probabilities is considered by many authors as a degenerative research programme [cf., e.g., Spohn (1981), p. 50], current *knowledge first* epistemology seems to be on a good way of forming a progressive one [cf., e.g., the overview provided in Feldbacher-Escamilla (2018), sect. 5]. An effective strategy of replying the critique consists in adopting also the methodology of *knowledge first* epistemology, which is *abductivism* in contrast to *deductivism* [cf. Williamson (2016)]. By this the problem of justifying (Pr1)–(Pr3) does not necessarily mean that one needs to provide a further argument or theory as, e.g., *belief first* epistemology does with the *action norm of belief* (Pr) discussed in section II. Rather, it suffices to simply point to the explanatory power of probabilism in order to justify it:

I see only very limited value in the project of arguing for the simple, plausible axioms of probability from more complex and no more plausible assumptions about the preferences of rational agents, for example, between dollar bets—money is not that basic [. . .]. An abductive methodology holds more promise: judge theories by their fruits [Williamson (2009), p. 333].

So, *knowledge first* can deal with the problem of providing an explanation for why we still act rational, although most of the time we do not act on knowledge. And it can circumvent the problem of providing a justification for the basic laws of graded belief, inasmuch as it employs an abductive methodology.

VI. CONCLUSION

In this paper we have given a detailed description of three approaches to rational action: The *belief first* approach, the approach of *logical probabilities*, and the *knowledge first* approach. The theory of logical probabilities states that we start with knowledge and try to derive by logical principles in a wide sense probabilistic information in form of degrees of belief. Given these degrees of belief we can use standard decision theoretic machinery in order to assess actions in terms of rationality. Now, although the theory of

logical probabilities sets high epistemic standards by putting knowledge first, it faces the problem that it provides no justification of the laws of probability. Several attempts to do so turned out to be more or less ad hoc and by this the programme is generally assessed as degenerative.

Some quantitative versions of the *belief first* approach to action resulted from a critique of logical probabilism: Adherents of this approach provide action norms of belief by help of betting-interpretations of degrees of belief and use degrees of belief in turn to formulate a belief norm of action. This allows for a justification of the laws of probability theory and explains how actions can be considered to be rational. Furthermore, due to its wide applicability it is also considered to be a highly progressive research programme. However, the epistemic standards it sets are lower and by this it cannot account for epistemic blamability, which expresses the fact that one can always be blamed for basing ones action not on knowledge, but on belief only.

Finally, *knowledge first* epistemology emerged from several deadlocks in epistemology that seem to be triggered by the *belief first* paradigm. By formulating a knowledge norm of action one can account for epistemic blamability by setting high epistemic standards for decisions, namely knowledge. The problem that we intuitively assess also actions as rational that are based on mere belief, but not knowledge, can be accounted for by differentiating between defective actions (which might be rational) and non-defective actions, i.e. actions based on knowledge (which are clearly rational). The *knowledge first* paradigm resulted from applying abductive methodology within philosophy, where it seems that some of the philosophical deadlocks can be overcome best by inverting the direction of explanatory priority from belief to knowledge. By applying the very same methodology to the problem of how to justify the laws of probability theory, the wide range of probabilism allows one to infer its validity without providing further arguments as, e.g. via action norms of belief. This is how *knowledge first* epistemology is putting knowledge first, while also providing a justification for (graded) belief.

In the following table the advantages and disadvantages of the approaches discussed in this paper are summarised:

Approach	Justifies <i>P_r</i> -laws	High standards	Blamability	Explains action	Is progressive
<i>Belief First</i>	✓	×	×	✓	✓
<i>Logical Probabilities</i>	×	✓	×	✓	×
<i>Knowledge First</i>	✓	✓	✓	✓	✓

It is clear that there are much more desiderata to be considered in order to assess a research programme as progressive or degenerative. Nevertheless, we hope to have indicated that *knowledge first* is able to provide an adequate norm for rational action. By this not only the *knowledge first* approach to belief, evidence, justification, and assertion seem to provide interesting alternatives to their traditional *belief first* counterparts. Also, its account of action seems to be feasible and worth further examination.*

Duesseldorf Center for Logic and Philosophy of Science (DCLPS)
University of Duesseldorf
Universitaetsstr. 1, 40225 Duesseldorf, Germany
E-mail: cj.feldbacher.escamilla@gmail.com

* This paper has been a finalist in the 2017 **teorema** Essay Prize for Young Scholars.

ACKNOWLEDGEMENTS

For valuable discussions on this topic I would like to thank Alexander Gebharter, Christoph Jäger, Mariana E. Martínez, and Gerhard Schurz. This research was supported by the German Research Foundation (DFG), research unit FOR 2495, A2 *Creative Abductive Inference*.

REFERENCES

- ARROW, K. J. (1963), *Social Choice and Individual Values*, 2nd Edition, Yale, Yale University Press.
- BENTON, M. A. (2014), “Knowledge Norms”; in *Internet Encyclopedia of Philosophy*, ed. by James Fieser and Bradley Dowden.
- BERTRAND, J. (1888)/(1907), *Calcul des Probabilités*, deuxième édition, Paris, Gauthier-Villars.
- BOONGALING, J. I. K. (2017), “Hintikka’s Socratic Epistemology Meets Gettier’s Counterexamples”; in *Kriterion – Journal of Philosophy* 31.3, pp. 25-56.
- BRAITHWAITE, R. B., ed. (1950), *The Foundations of Mathematics. And other Logical Essays*, with a Preface by G.E. Moore, London, Routledge & Kegan Paul LTD.
- (1926)/(1950), “Truth and Probability”, in Braithwaite (1950), pp. 156-198.
- (1929)/(1950), “Knowledge”, in Braithwaite (1950), pp. 258-259.
- CARNAP, R. (1962), *Logical Foundations of Probability*; London, Routledge and Kegan Paul.
- CARTER, J. A., GORDON, E. C., and JARVIS, B. (2017a), “Knowledge-First: An Introduction”; in Carter, Gordon, and Jarvis (2017b), pp. 1-18.

- eds. (2017b), *Knowledge-First: Approaches in Epistemology and Mind*, Oxford, Oxford University Press.
- DE FINETTI, B. (1972), *Probability, Induction and Statistics. The art of guessing*, Probability & Mathematical Statistics Series, John Wiley & Sons.
- FELDBACHER-ESCAMILLA, C. J. (2018), “Abductive Philosophy and Error”; in *Manuscript*.
- GAERTNER, W. C. (2009), *A Primer in Social Choice Theory*; Revised Edition, Oxford, Oxford University Press.
- GENEST, C. and ZIDEK, J. V. (1986-02), “Combining Probability Distributions: A Critique and an Annotated Bibliography”; in *Statistical Sciences* 1.1, pp. 114-135.
- GETTIER, E. L. (1963), “Is Justified True Belief Knowledge?”; in *Analysis* 23.6, pp. 121-123, url: <http://www.jstor.org/stable/3326922>.
- GREENOUGH, P. and PRITCHARD, D. eds. (2009), *Williamson on Knowledge*, Oxford, Oxford University Press.
- HÁJEK, A. (2005), “Scotching Dutch Books?”; in *Philosophical Perspectives* 19.1, pp. 139-151.
- HAWTHORNE, J. and STANLEY, J. (2008), “Knowledge and Action”; in *Journal of Philosophy* 105.10, pp. 571-590.
- JEFFREY, R. C. (1983), *The Logic of Decision*, Second Edition, Chicago, The University of Chicago Press.
- KAPLAN, M. (2009), “Williamson’s Casual Approach to Probabilism”; in *Williamson on Knowledge*, ed. by P. Greenough and D. Pritchard, Oxford, Oxford University Press, pp. 122-139.
- KEYNES, J. M. (1921), *A Treatise on Probability*; London, Macmillan and Co.
- LACKEY, J. (2010), “Acting on Knowledge”; in *Philosophical Perspectives* 24. 1, pp. 361-382, url: <http://dx.doi.org/10.1111/j.1520-8583.2010.00196.x>.
- MAHER, P. (2001), “Probabilities for Multiple Properties: The Models of Hesse and Carnap and Kemeny”; in *Erkenntnis* (1975-) 55.2, pp. 183-216, url: <http://www.jstor.org/stable/20013083>.
- MCGLYNN, A. (2014), *Knowledge First?*; New York, Palgrave Macmillan.
- PRITCHARD, D. (2014), “Knowledge Cannot be Lucky”; in *Contemporary Debates in Epistemology*, ed. by M. Steup, J. Turri, and E. Sosa, 2nd Edition, Chichester, John Wiley & Sons, pp. 152-164.
- SAVAGE, L. J. (1972), *The Foundations of Statistics*; Second Revised Edition, New York, Dover Publications.
- SCHURZ, G. (2013), *Philosophy of Science. A Unified Approach*, New York, Routledge.
- SPOHN, W. (1981), “Analogy and Inductive Logic: A Note on Niiniluoto”; in *Erkenntnis* (1975-) 16.1, pp. 35-52, url: <http://www.jstor.org/stable/20010716>.
- VAN FRAASSEN, Bas C. (1989), *Laws and Symmetry*; Oxford, Clarendon Press.
- VON NEUMANN, J. and MORGENSTERN, O. (1944)/(1953), *Theory of Games and Economic Behaviour*, Princeton, Princeton University Press.
- WILLIAMSON, T. (2000), *Knowledge and its Limits*; Oxford, Oxford University Press.

- (2009), “Replies to Critics. Reply to Mark Kaplan”; in Greenough and Pritchard (2009), pp. 333-340.
 - (2011), “Knowledge First Epistemology”; in *The Routledge Companion to Epistemology*, ed. by S. Bernecker and D. Pritchard, Routledge Philosophy Companions, London, Routledge, pp. 208-219.
 - (2016), “Abductive Philosophy”, in *The Philosophical Forum* 47.3-4, pp. 263-280, url: <http://dx.doi.org/10.1111/phil.12122>.
 - (2017), “Acting on Knowledge”; in Carter, Gordon, and Jarvis (2017b), pp. 163-181.
- ZAGZEBSKI, Linda (1994), “The Inescapability of Gettier Problems”; in *The Philosophical Quarterly* 44.174, pp. 65-73, url: <http://www.jstor.org/stable/2220147>.