Handelshochschule Leipzig (HHL)

On the Shape of Information Processing Functions

Hagen Lindstädt

HHL-Arbeitspapier Nr. 44

Abstract

The productivity of (human) information processing as an economic activity is a question raising some interest. Using Marschak's evaluation framework, Radner and Stiglitz have shown that under certain conditions the production function of this activity has increasing marginal returns in the beginning. This paper shows that under slightly different conditions this information processing function has repeated convexities with ongoing processing activity. Even for smooth changes in the signals' likelihoods, the function is only piecewise smooth with indifferentiable convexities at conditional changes of action. For linear likelihood functions the processing value shows to be piecewise linear with convexities at these levels.

Table of Contents

Abstract

The productivity of (human) information processing as an economic activity is a question raising some interest. Using Marschak's evaluation framework, Radner and Stiglitz have shown that under certain conditions the production function of this activity has increasing marginal returns in the beginning. This paper shows that under slightly different conditions this information processing function has repeated convexities with ongoing processing activity. Even for smooth changes in the signals' likelihoods, the function is only piecewise smooth with indifferentiable convexities at conditional changes of action. For linear likelihood functions the processing value shows to be piecewise linear with convexities at these levels.

1 Introduction

In today's economics, the fact that information is a major factor of production and of increasing importance as a source of competitive advantage is a generally accepted truism. It is only possible to evaluate information economically when a specific usage for the information is assumed. This usage is typically a certain decision problem, in which the information might change the decision makers' subjective prior probabilities for the environmental states. Furthermore, the evaluation has to be possible from a prior perspective before actually knowing the signal in order to be useful when deciding about acquiring and processing the information. The standard framework for this evaluation is briefly stated in section 2.

The analysis of the productivity of information as a factor of production and of information processing as an economic activity leads to a number of remarkable results. Just recalling a few, investments in information are often irreversible, what sometimes makes trading information difficult (Arrow 1974). Additionally, information does not disappear when used, so it can be used more than once, although its value might change while using it. Furthermore, information can be used for different tasks at the same time. These features demonstrate the fact that there are increasing marginal returns to information due to multiple usage.

Besides these considerations, Radner and Stiglitz (1984) have shown that under certain conditions there are increasing marginal returns to little information even if it is only used for one specific decision problem. As will be seen later, beginning information processing is not the only case for convexities when using information only once, i.e. in a single decision problem:

For slightly different conditions than those of Radner and Stiglitz, the activity of information acquisition and processing has repeated convexities even when continued.

2 Discussion of the Relevant Literature

2.1 Marschak's Value of Information and Blackwell's Condition on the informativeness of information structures

The standard in economic valuation of information is Marschak's model (Marschak 1954, 1959). It evaluates information with the gain a decision maker can extract from it in a specific decision problem due to a change in his subjective probabilities for the environmental states. For reasons of simplicity, we restrict ourselves to a risk neutral decision maker, a finite number of environmental states S_s and alternative courses of actions A_a to choose from. When deciding about acquiring costly information like a market report, the decision maker has to value the information on a prior basis, i.e. without knowing its content (what the market report actually says). This is achieved by distinguishing between the information I (the report) and the signals I_i, representing alternative versions of its content which the decision maker believes to be possible (e.g. growing, stagnating or shrinking revenue for a very simple market report).

The prior value of the information $I=(I_1,\ldots,I_I)$ before knowing the actual signal I_i is the expected additional payoff the decision maker can get from the decision due to possessing the information, i.e. knowing the signal. To calculate this payoff, he has to judge the likelihoods $p(I_i|S_s)$ for the signals conditional on the possible states and work out the conditional probabilities $p(S_s|I_i)$ using Bayes' well known equation. The resulting value of information equals the difference in expected payoff for the decision maker with and without the information (see (1)). The Markov matrix $(p(I_i|S_s))_{is}$ of likelihoods is called information structure. If U_{as} is the payoff of action A_a in state S_s in decision problem D and a* indexes the action with maximal expected payoff without the information, the value of information is like usually defined as

$$\mathrm{IV}_D\left(\left(\mathrm{p}(I_i|S_s)\right)_{is}\right)$$

$$:= \sum_{i=1}^{I} \mathrm{p}(I_i)\times\left[\max_{a=1,\ldots,A}\left\{\sum_{s=1}^{S}\mathrm{p}(S_s|I_i)\times U_{as}\right\}-\sum_{s=1}^{S}\mathrm{p}(S_s|I_i)\times U_{a^*s}\right]$$

$$
\begin{aligned}
&= \left[\sum_{i=1}^{I} \mathrm{p}(I_i) \times \max_{a=1,\dots,A} \left\{\sum_{s=1}^{S} \mathrm{p}(S_s \mid I_i) \times U_{as}\right\}\right] - \left[\sum_{i=1}^{I}\left(\sum_{s=1}^{S} \mathrm{p}(I_i) \times \mathrm{p}(S_s \mid I_i) \times U_{a^*s}\right)\right] \\
&= \left[\sum_{i=1}^{I} \mathrm{p}(I_i) \times \max_{a=1,\dots,A} \left\{\sum_{s=1}^{S} \mathrm{p}(S_s \mid I_i) \times U_{as}\right\}\right] - \left[\sum_{s=1}^{S}\left(\sum_{i=1}^{I} \mathrm{p}(S_s \mid I_i) \times \mathrm{p}(I_i)\right) \times U_{a^*s}\right] \\
&= \left[\sum_{i=1}^{I} \mathrm{p}(I_i) \times \max_{a=1,\dots,A} \left\{\sum_{s=1}^{S} \mathrm{p}(S_s \mid I_i) \times U_{as}\right\}\right] - \left[\sum_{s=1}^{S} \mathrm{p}(S_s) \times U_{a^*s}\right] \qquad (1)
\end{aligned}
$$

When evaluating information processing, the natural question arises how to compare two given information structures (matrices P and Q of likelihoods) with respect to their informativeness. In general, we know there is no way to measure the quantity of information as a real number, and of two given information structures neither one might be more informative. Still it is a well known result due to Blackwell (1951), that for any two given information structures P and Q the statement *"P is more informative than Q"* is equivalent to the existence of a matrix M such that

$$P \times M = Q \qquad (2)$$

The relation "more informative than" means that for any utility function and decision problem the value of information in P is at least equal to the value of information in Q.[1]

It seems natural to restrict using the term *information processing* to activities which lead to information structures becoming increasingly more informative. In other words, a matrix valued *information structure function*

$$\mathrm{IF}_P \;:\; t \rightarrow \left(\mathrm{p}_t\left(I_i \mid S_s\right)\right)_{is} \;=:\; P_t \qquad (3)$$

that maps real valued activity levels on information structures is only said to model information processing when for t $\geq$ r, P_t is more informative than P_r, that is when for each r and t with r $\leq$ t there exists an M_{rt} with

1 Blackwell (1951). Clearly, if such an M exists, Q can never be more informative, since Q can be constructed from P any time when knowing M. More surprising is the necessity for the existence of such an M. Notice that P and Q do not have to have the same number of signals since M is not restricted to have the same number of columns and rows. A more fundamental analysis of information structures is McGuire (1972).

$$\left(\pi_t\left(I_i \middle| S_s\right)\right)_{is} \times M_{rt} = \left(\pi_r\left(I_i \middle| S_s\right)\right)_{is} \tag{4}$$

2.2 The Radner / Stiglitz Result on the Nonconcavity of Starting Information Processing

When interpreting processing of information as an economic activity that may add value in a decision problem, a natural question about the input/output-relationship of this activity arises, i.e. about its production function. Here, the independent variable is the activity level of processing information, while the dependent variable is the value of the information processed. The decision problem D, the decision maker's preferences (here: risk neutral payoff maximizer), and a specific strategy for processing information have to be clear in order to make sense. With information processing strategy, we mean a plan or procedure about which "piece of information" is processed at which activity level. This production function with a real valued activity level t will be called *information processing function*.[2]

Radner and Stiglitz (1984) have shown a substantial and very interesting result for information processing functions with real valued arguments (activity levels) when processing information is costly: Under certain conditions that will be discussed below, the net marginal value of costly information processing is negative near the origin of zero activity, i.e. in a neighborhood of the origin. Radner and Stiglitz conclude that there have to be increasing marginal returns to information processing over some range of the parameter as soon as there is some amount of information processing with a positive net value.

Arrow (1985) has put this result into brief words: Given the Radner/Stiglitz conditions, "*a little information is never worth the cost*". Since already Arrow (1985) has directed attention towards the fact that the result strongly depends on its assumptions by giving examples that neither fulfill the conditions nor the result, it seems worthwhile to take a closer look at the four main assumptions from which Radner and Stiglitz were able to derive their strong result.[3]

---

[2] A formalization is given in (5).

[3] The assumptions not discussed are in only important to ensure that the Radner/Stiglitz-formalisms actually models costly information processing in a reasonable way.

Firstly, Radner and Stiglitz assume an infinite number of actions that can be indexed with a multidimensional, real valued parameter. An example for a decision problem with such an action set is the mixture of a securities portfolio. For k securities, the courses of action are the possible mixtures, a set that can be conveniently indexed with numbers from $[0;1]^k$ that add up to one – a (k-1)-dimensional subset of $[0;1]^k$.

They secondly assume payoff to be a continuous function in the parametrization of the actions, thus introducing a neighborhood structure between courses of action that apart from the parameters also applies to the payoff. Once a decision problem has an infinite number of alternative actions indexed by real valued parameters, it is natural to assume continuity: If any functional dependency between payoff and parametrization of alternative actions was allowed, one could hardly make statements about payoff shifts with changing probabilities and consequently about value of information, since slight changes in the optimal solution could easily imply a completely different payoff.

The third assumption is made on the informativeness of the information structure $(p_t(I_i|S_s))_{is}$ at t=0, i.e. before actually starting to process information: At t=0, the information structure is assumed to be *completely noninformative* for all signals I_i, noninformative meaning that the likelihoods $p_0(I_i|S_s)$ at t=0 are independent from s for all signals I_i. Like intended in the term, completely noninformative information structures bear no information that when knowing the signal could change the states' probabilities $p(S_s)$ or change the choice. Especially, the value of noninformative information structures (before costs) is always zero. So assuming noninformativeness means that information processing actually starts without information that is not already contained in the probabilities $p(S_s)$.

Accordingly, the fourth assumption rules out information processing procedures with sudden jumps or discontinuities in informativeness at the critical start: Radner and Stiglitz assume the information structure function $IF_P : t \rightarrow (p_t(I_i|S_s))_{is}$ to be differentiable in t=0. Consequently, there are no "sudden insights" of jumps in informativeness. Instead, information processing is assumed to happen smoothly. Both the third and fourth assumption seem to be a natural condition for deriving the convexity result, since for information processing procedures deviating from these assumptions, general results seem hard to obtain. This does, however, not mean that real world information processing might not in many cases proceed with sudden, non differentiable insights.

3 A Model for the Valuation of Information Processing

To model the production function of information processing, we map a real valued parameter representing the information processing activity to its value, i.e. to the information structure that is connected with that activity. As a restriction, we only allow information structures becoming successively more informative, i.e. we restrict processing to information structure functions $IF_P : t \rightarrow (p_t(I_i|S_s))_{is}$ like in equation (3) that fulfill (4). Additionally, IF_P is assumed to be continuous in every component, i.e. the likelihood functions $IF_{P;is} : t \rightarrow p_t(I_i|S_s)$ are all continuous.

So like in Radner and Stiglitz (1984) with a slightly different interpretation, the information processing activity is real valued. Diverging from their approach we shall restrict ourselves to decision problems with a finite number of alternative courses of action A_a, environmental states S_s, information signals I_i, and a risk neutral decision maker. If information processing starts with zero information (completely noninformative information structure), the likelihood $p_0(I_i|S_s)$ for all signals I_i is independent of the state S_s. Unlike in the Radner / Stiglitz model, this is however not required.

The signals represent possible outcomes of information processing from a prior perspective. With growing information processing activity t, the information structures attached become more informative, meaning the stochastic indication of the signals towards the states gets increasingly stronger. Eventually but not necessarily, the point of complete information is reached, where the conditional probability $p_t(S_s|I_i)$ for all states and signals is either 1 or 0. Differing from evaluating processing of information in one single step, the information here is processed continuously. The signals I_i represent prior possibilities for processing results. It should be noted that for reasons of simplicity, the set of possible signals does not depend on the processing activity t.

It must also be kept in mind that the value of the information structures is always calculated from the prior perspective, implying that a decision about an optimal information processing activity derived from the model uses only the information given before processing starts: the decision whether to stop or to proceed with information processing here is not adaptive in the sense that it uses the signals being processed along the way. This is important since an adaptive strategy might lead to better results when feasible: with the constantly changing base of the decision maker's information while processing, the relevant adaptive information processing function for the remaining processing activity changes steadily.

Since the value of a given information structure (likelihood matrix) $(q(I_i|S_s))_{is}$ is calculated according to (1), and (3) gives the continuous sequence of information structures satisfying (4), the information processing function IPF can be defined according to

$$\mathrm{IPF}_{D;P}(t) \quad = \quad \mathrm{IV}_D\left(\mathrm{IF}_P(t)\right) \tag{5}$$

IPF is the "production function" for the information processing activity, mapping each processing activity t on the prior value of its associated information structure before cost. The information structure function P_t models information processing, i.e. by assumption condition (4) is fulfilled for all $r \leq t$. Hence, it is immediately clear from Blackwell's result that $\mathrm{IPF}_{D;P}$ is monotonously non decreasing in t.

For every processing activity t and signal I_i, the action with the maximal expected payoff U will be marked with the index $a^*_{i,t}$. In the case of zero information at t=0, clearly $a^* = a^*_{i,0}$ is the optimal action independently of signal I_i without additional information processing. Starting from (1) and (5) using this notation, we obtain

$$\begin{aligned}
\mathrm{IPF}_{D;P}(t) &= \sum_{i=1}^{I} \mathrm{p}_t(I_i) \times \left[\max_{a=1,\ldots A} \left\{ \sum_{s=1}^{S} \mathrm{p}_t(S_s|I_i) \times U_{as} \right\} - \sum_{s=1}^{S} \mathrm{p}_t(S_s|I_i) \times U_{a^*s} \right] \\
&= \sum_{i=1}^{I} \mathrm{p}_t(I_i) \times \left[\sum_{s=1}^{S} \mathrm{p}_t(S_s|I_i) \times \left(U_{a^*_{i;t}s} - U_{a^*s} \right) \right] \\
&= \left[\sum_{i=1}^{I} \sum_{s=1}^{S} \mathrm{p}_t(I_i) \times \frac{\mathrm{p}_t(I_i|S_s) \times \mathrm{p}(S_s)}{\mathrm{p}_t(I_i)} \times \left(U_{a^*_{i;t}s} - U_{a^*s} \right) \right] \\
&= \sum_{i=1}^{I} \sum_{s=1}^{S} \mathrm{p}(S_s) \times \left(U_{a^*_{i;t}s} - U_{a^*s} \right) \times \mathrm{p}_t(I_i|S_s)
\end{aligned} \tag{6}$$

Discussion: A closer look at the last term shows an important feature of the information processing function. If for a given signal I_i and activity level t* there is a single action with maximal expected payoff, implying that no two or more alternative courses of action have the same expected payoff at (t*,i), then there will be a neighborhood of t* for which this action stays optimal for I_i, because IF_P was assumed to be continuous in every component $p_t(I_i|S_s)$.

Case 1: If this neighborhood can at t* be found for every signal I_i, then there will also be a neighborhood near t* in which the actions being optimal for the respective signals I_i stay the same. In that case, the payoff difference in the last term of (6) is constant within that neighborhood, i.e. it does not depend on t. Since also the prior state probabilities $p(S_s)$ do not depend on t, the information processing function within that neighborhood is a linear mixture of the likelihood functions' components $p_t(I_i|S_s)$.

Case 2: If there is a signal I_i for which in t* at least two actions yield the same expected payoff, then for this signal the optimal action changes at t*. For these t* a change of the optimal course of action happens conditionally on a signal I_i.

In section 4, specific functions IF_P will be analyzed for their associated information processing functions $IPF_{D;\,P}$.

4 Results: The Shape of Information Processing Functions

It is clear that even if all likelihoods $p_t(S_s|I_i)$ differentiably depend on t, the same cannot be said for the information processing function. Formally, this is caused by the maximum function in the value of information.

Result 1: If information processing is monotonously non decreasing and smooth in the sense that all likelihoods differentiably depend on t, then the information processing function is continuous everywhere and piecewise differentiable with isolated activity levels where it can usually not be differentiated. These t correspond to processing activity levels for which there is a change in the optimal action for certain signals I_i (*"conditional change of action"*). At these indifferentiable levels the function is convex even for concave likelihoods, i.e. the marginal return on its right are larger than those on its left.

The continuity results directly from the continuity of all likelihood functions $p_t(I_i|S_s)$. In *case 1* of the brief discussion following (6), if all $p_t(I_i|S_s)$ are differentiable in such a t*, the same is true for $IPF_{D;P}$ at this particular t* and within its neighborhood where the optimal action does not change for any signal I_i. In *case 2* of that discussion, there is a conditional change in the optimal action for at least one signal I_i at t*. At these t*, the payoff optimal action and thus the payoff difference in (6) changes for at least one signal. The information processing function usually cannot be differentiated in such t*, but clearly is still continuous.

The growth in slope follows looking at the last term in (6): If the function cannot be differentiated at t* as described, then for at least one signal I_i there are two courses of action yielding the same expected payoff at t*, and for t>t* a different action becomes conditionally optimal under I_i in expected payoff. If the decision maker wouldn't change his course of action for t>t* under this I_i, his information processing function $IPF_{D;P}$ would suboptimally stay at the previous non decreasing path (see (6)). Changing to the conditionally superior action for t>t* at I_i changes the "mixing factors" for the likelihood functions $p_t(I_i|S_s)$ of $IPF_{D;P}$.[4]

With the model described in section 3, it is easy to see that starting at t=0 with completely noninformative information and a single best course of action A_{a*}, a result similar to Radner and Stiglitz applies: At the start, information processing is completely worthless, since the alternative courses of action have to compensate continuously in t for their disadvantage in expected payoff.

Result 2: In the model described in section 3, zero information at t=0, and a single best action A_{a*}, i.e. an unique action with maximal expected payoff, information processing is completely worthless at the beginning: In this case, there exists an t*>0 such that for all $t \leq t^*$: $IPF_{D;\,P}(t) = 0$.

This result follows directly from (6) and the reasoning at the end of the previous section *(case 1)*. If there is zero information at t=0, then the $p_0(I_i|S_s)$ are independent of the states S_s, hence $p_0(S_s|I_i)=p(S_s)$ for all signals I_i. If there is an unique action with maximal expected payoff, the payoff of the other possible courses of action have to overcome a difference in expected payoff before they can be optimal for certain signals I_i. Since it is assumed that all $p_t(I_i|S_s)$ change continuously with t, the same is true for $p_t(S_s|I_i)$. Hence under each signal, A_{a*} stays optimal for small enough changes of t. Thus taking the minimal one of these small enough changes for all signals, we have

---

4 Schauenberg (1986) has demonstrated this result for the special case of two environmental states.

found a t* up to which $IPF_{D;P}$ stays at zero and the processing is of no value (see (6),(1)).

Example 1

A risk neutral decision maker faces a decision problem with six possible courses of action and two environmental states with attached prior probabilities and information structures arising during processing (see table 1).

Table 1: Decision Problem D and Information Structure Function P_t[5]

D	**S_1**	**S_2**	
$p(S_s)$	**0,5**	**0,5**	**E_U**
A_1	220.00	-255.00	-17.50
A_2	150.00	25.00	87.50
A_3	100.00	100.00	100.00
A_4	63.33	130.00	96.67
A_5	-6.67	160.00	76.67
A_6	-366.67	200.00	-83.33

P_t

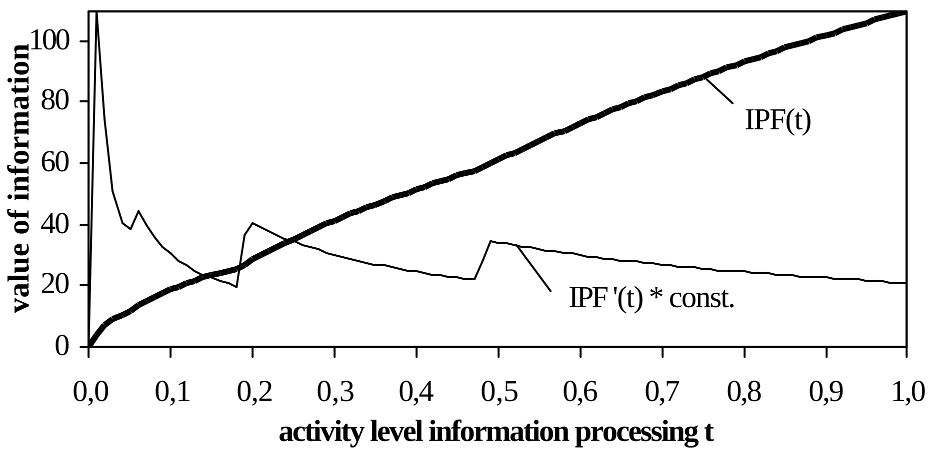
$p_t(I_i\|S_s)$	**I_1**	**I_2**	**$p(S_s)$**
S_1	$½+½t^{0.3}$	$½-½t^{0.3}$	0,5
S_2	$½-½t^{0.3}$	$½+½t^{0.3}$	0,5
$p_t(I_i)$	0,5	0,5	

Figure 1 shows the information processing function and its derivative for this case of differentiable information processing and zero information at activity level t=0.[6] Clearly, the value of perfect information and thus an upper bound for the function is 110. In this special case, $p_t(I_i)$ is independent from t, since $p(S_1)=p(S_2)$.

5 In interpreting t^a as learning pace, it helps to know that in asymptotic statistics the quality of statistical results usually improves with a speed of $\sqrt{n}$ on average (a=0.5), where n is the discrete number of independent trials, see Rüschendorf (1988).

6 For visual reasons the derivative is multiplied with a constant factor.

Figure 1: Information Processing Function for the decision problem and the information structure function in tables 1 and 2



It seems reasonable to look at information processing functions of simple likelihood functions $p_t(I_i|S_s)$ and hence IF_P. The simplest interesting case is a constant change in all likelihoods $p_t(I_i|S_s)$, i.e. the case of linear likelihood functions in each component: We will call information processing linear and the associated learning having a *constant pace* when for all states S_s and all signals I_i there are real parameters a_{is} and b_{is} such that

$$\mathrm{p}_t\left(I_i \,|\, S_s\right) \quad = \quad a_{is} \times t \, + \, b_{is} \tag{7}$$

If the absolute value of the a_{is} is comparatively high for all signals and states, the likelihoods change relatively fast with t and learning has a high pace and vice versa. It is natural to formalize the pace of learning at t as sum of the likelihood's differentials at t in absolute values:

$$\mathrm{LP}_P(I;t) \quad = \quad \sum_{i=1}^{I} \sum_{s=1}^{S} \left| \frac{\partial \mathrm{p}_t(I_i|S_s)}{\partial t} \right| \tag{8}$$

With this definition, linear information processing as defined in (7) has constant learning pace at all t.

Result 3: For linear information processing, the information processing function is piecewise linear between the activity levels where it cannot be differentiated. At these levels, its slope grows.

The piecewise linearity between the conditional changes of action (result 1) follows directly from (6) and the considerations thereafter, together with all likelihood functions being linear (see (7)). The convexity at the activity levels where $IPF_{D;P}$ cannot be differentiated is a special case of result 1.

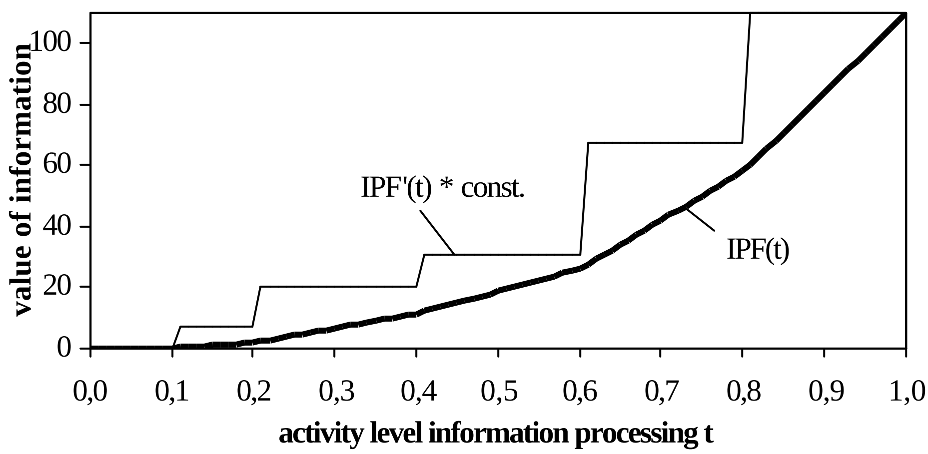
Example 2

In the decision problem discussed in example 1 (see table 1), the risk neutral decision maker now faces an information processing opportunity as described in table 2. The information processing function in this case of linear information processing is shown in figure 2. At t=0 there is again zero information.

Table 2: Information Structure Function for Linear Information Processing

$p_t(I_i\|S_s)$	I_1	I_2	$p(S_s)$
S_1	½+½t	½-½t	0,5
S_2	½-½t	½+½t	0,5
$p_t(I_i)$	0,5	0,5	

Figure 2: Information Processing Function for the Decision Problem D (table 1) and Linear Information Processing (IF in table 2)



5 Summary and Outlook

It has been shown that increasing marginal returns to information processing even for single usage are not restricted to the start of the processing activity, but do also apply repeatedly through its course. The value of processing information is a locally linear combination of the associated likelihood functions with the payoff of the respective optimal actions serving as weighting factors. Especially, if all likelihood functions are linear, information processing yields piecewise linear expected payoffs with increasing slope at the activity levels with conditional change in optimal action for certain signals.

There are many consequences of the convexities in information processing functions. Firstly, for finding the optimal (prior) degree of information processing, the standard marginality criterion for optimality does not apply. Seemingly innocent statements like March (1994, p.25), that *"rational decision makers can be expected to invest in information up to the point at which expected cost equals the marginal expected return"* prove to be wrong in this generality, since the productivity of information processing systematically has increasing marginal returns: Even if marginal returns and (constant) marginal costs equal each other, it might well be useful to continue processing if such a convexity overcompensating a marginal loss is close. In order to make this argument, one does not have to rely on repeated use of information.

Secondly, there are major implications for issues of coordination, control and optimal size of the firm. Just to name two important ones, already Arrow (1975) discussed the positive implications for vertical integration due to uncertainty reduction. Similarly, Radner (1992) discusses whether possible disadvantages of size that manifest in over proportional growth in coordination cost can be more than compensated by increasing marginal returns of information.

A third area where important implications can certainly be expected is the organization of knowledge intense industries and the creation of competitive advantage in these. After analyzing situations and implications of increasing marginal returns, Arthur (1996) provided a management level overview of such considerations. These and other important implications are largely subject to further research and will not be further discussed here.

References

Arrow, K.J. (1974): *Limits of Organization*. New York 1974.

Arrow, K.J. (1975): "Vertical Integration and Communication." Bell Journal of Economics 6 (1975), p.173-183.

Arrow, K.J. (1985): "Informational Structure of the Firm." American Economic Review, Papers and Proceedings, 75 (1985), p.303-307.

Arthur, W.B. (1996): "Increasing Returns and the New World of Business." Harvard Business Review 1996, p.100-109.

Blackwell, D. (1951): "Equivalent Comparison of Experiments." Annals of Mathematical Statistics 24 (1953), p.265-272.

March, J.G. (1994): *A Primer on Decision Making*. New York 1994.

Marschak, J. (1954): *Toward an Economic Theory of Organization and Information*. In: Decision Processes. Thrall, R.M., C.H. Coombs, and R.L. Davis (eds.), New York 1954, p.187-220.

Marschak, J. (1959): *Efficient and Viable Organization Forms*. In: Modern Organization Theory. Haire, M. (ed.), New York 1959, p.307-320.

McGuire, C.B. (1972): *Comparison of Information Structures*. In: Decision and Organization. McGuire, C.B. and R. Radner (eds.), 2nd edition New York 1986, p.101-130.

Radner, R. (1992): "Hierarchy: The Economics of Managing." Journal of Economic Literature 30 (1992), p.1382-1415.

Radner, R. and J.E. Stiglitz (1984): *A Nonconcavity in the Value of Information*. In: Bayesian Models in Economic Theory. Boyer, M. and R.E. Kihlstrom (eds.), Amsterdam 1984, p.33-52.

Rüschendorf, L. (1988): *Asymptotische Statistik*. Stuttgart 1988.

Schauenberg, B. (1986): *Der Verlauf von Informationswertfunktionen*. In: Information und Wirtschaftlichkeit. Ballwieser, W. and K.-H. Berger (eds.), Wiesbaden 1986, p.229-251.