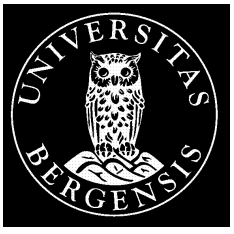
WORKING PAPERS IN ECONOMICS

No. 6/13

JOACHIM RAMM, SIGVE TJØTTA AND GAUTE TORSVIK

INCENTIVES AND CREATIVITY IN GROUPS



Department of Economics
UNIVERSITY OF BERGEN

Incentives and creativity in groups*

Joachim Ramm†, Sigve Tjøtta‡and Gaute Torsvik§

June, 2013

Abstract

It has been argued that monetary incentives restrain individual creativity and hamper performance in jobs requiring out of the box thinking. This paper reports from an experiment designed to test if the negative incentive effect is present also when individuals work together to solve such problems. We do not find a negative impact of incentives on group performance. As a comparison we ran the same experiment (the Candle Problem) with and without incentives for individuals as well. Incentives did not reduce performance there either. Comparing individuals with groups we find that team-work facilitates creative problems solving. Individuals appear to be more creative when working together than when working alone.

JEL Codes: J24, J31, O31, M11.

Keywords: Incentives, innovation, creativity.

*We thank Kjell Arne Brekke, Kalle Moene, Raquel Fernandez and seminar participants at the Choice Lab seminar (NHH) and at the Department of Economics, University of Bergen, for helpful comments. We thank Sebastian Skancke and Eirik Strømland for excellent research assistance.

†Minestry of Local Governemnet and Regional Development

‡University of Bergen, sigve.tjotta@econ.uib.no

§University of Bergen, gaute.torsvik@econ.uib.no

1 Introduction

Some tasks can be solved in a mechanical way by exploiting existing knowledge and by following well rehearsed routines and practices. Other problems require a more novel and explorative approach. Finding a solution often entails a reframing and restructuring of the task. One has to look at the problem from an unusual angle - "think outside the box" - to make progress. How do monetary rewards affect our ability and effort to solve such insight problems?

In his best selling book *Drive,* Dan Pink (Pink (2010)) hammers the idea that monetary incentives impair individual creativity. As evidence, he refers to results from experiments done in psychology and economics, such as the "Candle Problem". This experiment was invented by Karl Duncker (Duncker (1945)) to study "functional fixedness" in cognitive reasoning. Duncker showed that it is hard for grown ups to solve problems requiring out of the box thinking; adults have acquired a mental block against using objects in an unfamiliar, novel way. The psychologist Sam Glucksberg (Glucksberg (1962)) ran a Candle Problem experiment with monetary rewards and found that incentives aggravate the problem of functional fixedness.

Creative work is often done in groups, it is therefore important to find out whether the negative incentive effect is present also when individuals work together to solve insight problems. It could be that the challenge and intensity of operating in a group, playing ideas with team-mates, overturns the mechanisms that induce large stakes to increase functional fixedness (less creativity) for individuals. Maybe explorative work is organized in groups partly because the rewards that are awaiting those who find a solution, do not block creativity in teams.

To examine the role of monetary incentives in groups, we ran a Candle Problem experiment where two individuals worked together. We contrasted the outcome of an incentivized trial, where participants received a prize if they performed well, with the outcome in a non-incentivized session. The results show no negative effect of incentives. In fact average performance is slightly better when performance is incentivized, but the difference is not statistically significant.

Next, we ran the same experiment with individual participants. Based on the previous finding we expected incentives to have a negative effect on performance, but again there was no difference in performance between incentivized and non-incentivized participants. We failed to replicate the results in Glucksberg (1962).

We think the fact that we cannot reject the null hypothesis warrant some attention as it goes counter to the widespread notion that incentivizing performance impair the

creativity needed to solve the Candle Problem. In addition to this "null result" we find an interesting difference between individuals and groups.

Groups perform much better in the Candle Problem. Of course just the fact that a group consists of several minds - two in our case - implies that we should expect them to arrive at a solution faster than a single mind. In their survey of experimental evidence on decision making, Charness and Sutter (2012) note that if each person in a population has a probability p of having the insight that solves the problem, then a group of n randomly sampled individuals has a likelihood $1-(1-p)^n$ for reaching a solution, which is lager than p for $p<1$. They refer to $1-(1-p)^n$ as the *truth wins benchmark* and note that although groups typically do better than individuals in insight problems, they rarely reach the *truth wins benchmark*. In our experiment groups exceed this benchmark. Individuals appear to become more creative when working together than when working alone.

Despite a booming research on group effects in decision making we are not aware of any studies examining group effects on creative problem solving.[1] Hence, we believe this paper provides novel insights into an important question.

2 The research questions

We address two questions. How do groups solving insight problems respond to monetary incentives? And irrespective of the extrinsic rewards that are awaiting, are groups better at solving insight problems than individuals?

There are two important differences between mechanical tasks and insight problems. First, experience is not always an advantage for solving insight problems as "functional fixedness" - the inability to consider an object or a problem from an unusual angle - may prevent the kind of exploration that is needed to find a solution (Duncker 1945). Another difference is that individuals usually have a higher level of intrinsic motivation for insight problems than for mechanical problems (Deci 1971).

Due to these differences extrinsic rewards - performance pay - may work differently on insight problems and mechanical problems. It is uncontroversial - and well documented - that sufficiently strong performance incentives elevate the attention and effort of those solving mechanic tasks.[2] This may not be true for insight problems. In a well

[1] Charness and Sutter (2012) and Kugler et al. (2012) survey recent experimental work on team decisions. Charness and Sutter conclude that teams tend to behave more in line with game theoretic predictions than individuals do. Kugler et al argue that heterogeneity in the levels of cognitive abilities, beliefs and attitudes towards risk among team members is causal for the difference in decision making by groups and individuals.

[2] Performance pay may, however, not always work as intended (by the incentive providers) for more

Standard setup

Solution

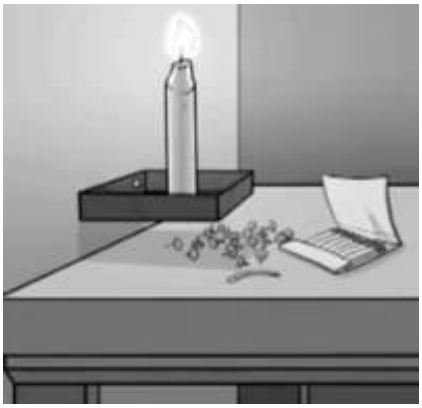


Figure 1. The Candle Problem

known study Edward L. Deci argues that for tasks that are high on intrinsic motivation, extrinsic rewards may lead to over-justification and may crowd out the inner drive to such an extent that the incentives reduces performance Deci (1971).[3]

In a study that has received less attention - at least until Dan Pink revitalized it - Sam Glucksberg suggests that extrinsic incentives will aggravate functional fixedness; incentives may prevent exploration and out of the box thinking. To test this hypothesis Glucksberg ran a variant of the Candle Problem. It is this problem we extend by letting groups solve it.

In the Candle Problem participants are presented with a candle, a box containing tacks and a book of matches. They are asked to attach the candle to the wall by only using the objects presented so that the candle can burn properly, and that no wax will drip on the table or on the floor. The participants usually have a time limit to find a solution. To solve this problem one has get the insight that the box containing the tacks can be used as a platform for the candle. One literally has to engage in out of the box thinking. See figure 1 for the standard setup and solution to the Candle Problem.

Note that if the tacks are delivered outside the box, the issue of functional fixedness is not as pressing (when tacks are in the box one has to reinvent the box as a platform). This fact was exploited by Glucksberg when he examined the effect of economic

mechanical tasks either, as incentives may lead to task shifting and gaming (Holmstrom and Milgrom, 1991, Oyer, 1998) and may also displace other forms of extrinsic motivations (fairness concerns, social appraisal, etc.) (Gneezy and Rustichini, 2000; Bowles and Polania-Reyes, 2012)

[3]Deci ran an experiment and found a pattern in performance consistent with then notion that incentives may have a detrimental effect on puzzle solving. The interpretation of this result - that it leads to over justification which then crowds out intrinsic motivation - is contested (Fehr and Falk, 2002). There are however several other studies that show a similar effect, see Frey and Jegen (2001) for an overview.

incentives.

In Glucksberg's experiment, participants were divided into two groups. Each group conducted a different variation of the "Candle Problem," one in which the box was empty and the tacks were placed on the table ("tacks outside box" version), and one in which the tacks were situated inside the box ("tacks inside box" version).

The participants in both groups were assigned either to a non-reward condition named "low-drive" (non- incentivized group), or a reward condition where they could receive \$5 for being among the top 25% fastest solvers, or \$20 for being the fastest, named the "high-drive" condition (incentivized group).

A failure was defined as being unable to find the correct solution within the time limit of 15 minutes. In the "tacks outside box" version, the results were as expected; the group with the financial incentive performed better, they had a lower fraction of failures and a shorter solution time. The surprise came in the "tacks inside the box" version, this time the incentivized group performed significantly *worse* than the group working without incentives.

Our first research question is to examine if the - apparent - negative incentive effect on human creativity carries over to groups? It seems like an important research question as creativity and novel solutions are often associated with considerable rewards. The rewards can be part of a designed incentives scheme within an organization but that is not always the case. The rewards can also come from higher market value, improved career opportunities and social appraisal. If high stakes impair individual creativity by preventing out of the box thinking it is of considerable interest to examine if this also is the case when a group of individuals interact to solve such problems.

Another, perhaps more basic research question is if there is a positive group effect for these kinds of problems. Suppose you have 10 individuals available and you want to minimize the expected time it takes to solve an insight problem (the Candle Problem for example), should you arrange 5 groups of two individuals or let the 10 persons grope for a solution in isolation? The answer to this problem is not obvious. It has been shown in both strategic and non-strategic settings that it matters whether the decision maker is an individual or a group of individuals. In general, groups appear to be more strategic and goal oriented decision makers. They appear to be better at processing information and they solve general reasoning problems faster than individuals. [4] These findings are suggestive for what we should expect in our case as well, but it has also been shown that groups do not pool information in an efficient way and that some perspectives,

[4]Confer Charness and Sutter (2012) and Kugler, Kausel and Kocher (2012) provide recent surveys on how team decisions compare with individual decisions.

some ideas, may be suppressed by the majority or by a subgroup of dominating persons within a group ((Kugler et al., 2012)) . Hence it could be that fewer ideas will be explored when individuals operate in a group than if the members groped for a solution in isolation.[5]

3 The Experiments

As our goal is to check if money rewards hamper creativity also in groups, we only organized the Candle Problem with stacks inside the box. [6]

3.1 The group session

In our first session (May 2012) we recruited 60 undergraduate students from the University of Bergen. We randomly grouped them together two and two and - again randomly - assigned pairs either to the incentivized or the non-incentivized version of the experiment.

All participants received 100 NOK upon arrival, 1 NOK is approximately 1/6 US dollar. Those in the incentivized version of the experiment were informed that each member of the pair with the shortest solution time would receive a prize of 1000 NOK, pairs in the second and third place would each get 200 NOK. They where also told that that the total available time was 15 min.

Participants in the session without incentives where told that the task was to solve the problem and that time would be measured and that there was a time cap on 15 min.

We ended up with 16 incentivized and 14 non-incentivized pairs. After completing the task, the participants were asked if they were familiar with the problem. Those who were are excluded from the analysis. Our analysis is based on 15 incentivized and 10 non-incentivized pairs.

[5]We have never seen a group version of the Candle Problem (with or without incentives). There are however many other hypothesis that has been examined by using the Candle Problem: Does living abroad increase creativity Maddox and Galinsky (2009): Are older kids more prone to functional fixedness than younger ones, German and Defeyter (2000): Are creative individuals more dishonest Gino et al. (2012) to name a few.

[6]Due to the difficulty of finding a book of matches - used in the Glucksberg experiment - the participants got a box of matches.

3.2 Individual sessions

To contrast our finding in the group session with the outcome when individuals solve the problem in isolation - that is, to replication the Glucksberg study - we requited (a month later, june 2012) 29 students from University of Bergen; 14 students solved an individual incentivized Candle Problem and 15 participated in a session without incentives. We ran another individual session in February 2013. We recruited 50 students to participate in the experiment in order to have a sample large enough to check for gender differences in how participants responded to incentives (in the Glucksberg experiment only men participated). The individual sessions are pooled together when we compare individuals and groups.

Participants were randomly assigned to the different treatments (incentives or not). The magnitude of the individual rewards were also the same as in the group session. We asked participants in the individual sessions if they were familiar with the Candle Problem. Only 3 confirmed (all in February 2013 sessions), and they are excluded from our analysis which is based on 39 incentivized and 40 non-incentivized individuals

4 Results

We compare two performance measures across sessions, success ratios i.e. the fraction of participants that manage to solve the problem, and for the successful, the time they use to solve the problem.

Incentives and creativity

Our first finding is that monetary incentive does not impact on the two above mentioned performance indicators, neither for individuals nor for groups. The mean solution times together with the success ratios are presented in Table 1.

Table 1. Performance with and without incentives

		Success ratio	Solution Time
		Fraction (N)	Mean (Stdv)
Individuals	Non-incentivized	0.77 (40)	5.38 (3.65)
	Incentivized	0.72 (39)	6.45 (3.90)
	All	0.75 (79)	5.89 (3.78)
Groups	Non-incentivized	1.00 (10)	3.41 (1.55)
	Incentivized	1.00 (15)	2.97 (2.07)
	All	1.00 (25)	3.15 (2.69)

The success ratio is given by the number of individuals (pairs) that solved the problem within 15 minutes divided by the number that participated in the session. The mean solution time is conditional on having solved the problem.

All groups solved the candle light problem within the time limit. Groups that are incentivized solve the problem slightly faster than the non-incentivized groups, but the difference in average solving time is not significant, according to a Kruskal-Wallis rank test, $p = 0.36$, two sided. Incentivized individuals have a slightly lower success rate than those who perform without money rewards, but, again, the difference is not significant (a binomial probability test, $p = 0.49$, two sided test). Conditional on having succeeded, individuals who are incentivized need more time to solve the problem than those who are not rewarded. The difference is relatively small and insignificant (Kruskal-Wallis rank test, $p = 0.26$, two sided).

We are not able to reproduce the Glucksberg (1962) results. This could be due to the gender composition of the participants. We use a mix of male and female students, whereas in the Glucksberg experiment only male students solved the problem. To examine this possibility we split our individual sessions into a male and a female subsample. There is, however, no significant gender difference in how individuals respond to incentives.[7] The only statistically significant gender difference we find is that a higher

[7]Female success ratios are 0.64 and 0.68 in the incentivized and non-incentivized sessions. They are not significantly different according to a binomial probability test ($p = 0.81$, two sided). On average incentivized females need more time to solve the problem than those who work without extrinsic rewards, 6.72 minutes compared to 4.81 minutes. The difference is significant at moderate significance level (Kruskal-Wallis rank test, $p = 0.09$, two sided). Male success ratios are 0.82 and 0.86 for incentivized and non-incentivized respectively, the difference is not significant different (a binomial probability test, $p = 0.72$, two sided). Mean solution times are 6.18 minutes incentivized men compared to 5.78 minutes for incentivized. Again the difference is not significant (Kruskal-Wallis rank test, $p = 0.89$, two sided).

fraction of men manage to solve the problem (a binomial probability test, two-sided, gives $p = 0.04$). Solution times are not significantly different for males and females.

Group creativity

Table 1 indicate that there is a positive group effect. Groups perform considerable better than individuals, both in terms of success ratios and mean solution times. All groups succeeded in solving the candle light problem compared with individuals success ratios of 0.72 and 0.77 for incentivized and non-incentivized, respectively. The solution time for groups are significantly lower than individuals both for incentivized and non-incentivized ($p = 0.002$ and $p = 0.08$ respectively, Kruskal-Wallis rank tests, one sided test).

It is not surprising that a pair of individuals solve an insight problem faster than one individual. A pair solve the problem as soon as one of the members has the "insight". So even in the absence of any synergies one would expect teams to outperform individuals in the Candle Problem. In order to check if there is a profound team effect, we construct pairs from the individual session and compare their performance with the performance in the real group session. The performance of the simulated groups is recorded as follows. First, we draw, without replacement, a random sample of m pairs from the individual sessions.[8] The performance of a simulated pair is given by the best performance of the two individuals comprising a specific pair. If both individuals managed to solve the candle light problem, this constructed group is successful and its solution time is the fastest individual solution time of pair. If only one in the constructed pair managed to solve the probe in the individual session the group is successful and its solution time is the successful partner's time. If neither of the individuals in the constructed pair managed to solve the problem in the individual session, the pair of them is recorded as unsuccessful. In the third step, we calculate the mean performance for these m constructed pairs. Finally, we ran a Monte Carlo simulation (1000 simulations) and we compare the mean performance in our true group session with the distribution of mean performances of the m simulated pairs.

If there is just a mechanical group effect the mean performance in our true group session ought to be located close to the center of the simulated distribution of means. Table 2 summarizes the results.

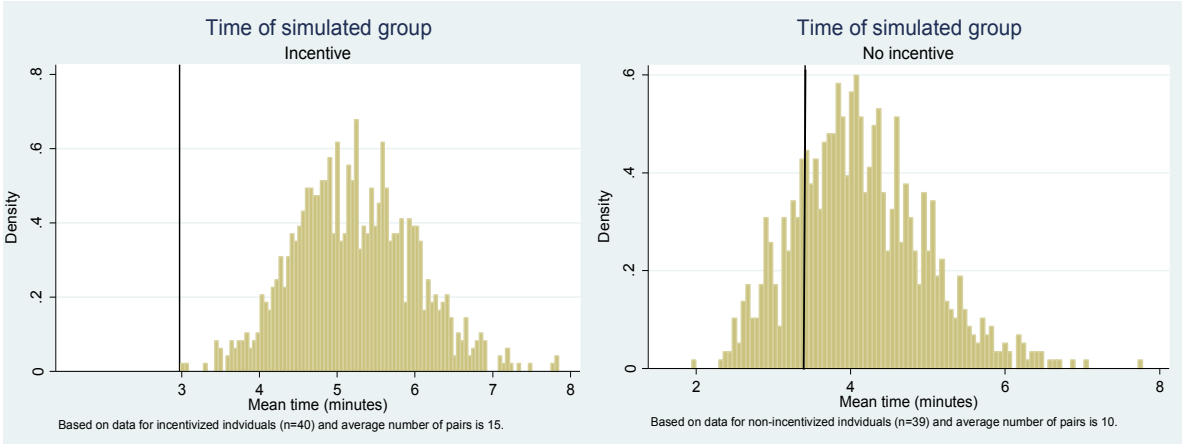
[8]Were m is equal to the number of pairs in the group session we compare with.

Table 2. Comparing simulated groups with real groups

	Incentive		All
	No	Yes	
Success ratio in sessions with simulated pairs (Stdv)	0.950 (0.062)	0.921 (0.060)	0.942 (0,.040)
Simulated pair sessions with success rate $\geq$ than the success rate in true group	550/1000	219/1000	173/1000
Mean solution time in sessions with simulated pairs (Stdv)	4.12 (0.84)	5.22 (0.76)	4.36 (0.50)
Simulated pair sessions with mean solution time $\leq$ than the mean solution time in true group	201/1000	0/1000	3/1000

The mean success ratio for the constructed groups is 0.942 (standard deviation of 0.040) obviously less than the observed overall success ratio of 1. But the difference does not seem to be significantly as 173 of 1000 of constructed group samples also have a success ratio of 1. Splitting the constructed pairs into two subsamples depending on whether or not they are incentivized does not change this result. We also calculated the *truth wins benchmark* based on the success ratios for individual sessions; 0.938 for the whole sample which is close to mean success ratios for simulated group.

The real groups solve the candle light problem faster than the constructed groups. The mean solution time is 3.15 minutes for the true groups compared to 4.36 minutes for the constructed groups. Only 3 out of 1000 mean solutions time of constructed sample are less or equal to mean solution time of the true groups. It is especially among the incentivized sessions that the differences between the simulated and the real groups are large. This is illustrated in Figure 1.



The black vertical line indicates the mean solution time in the real group session

Figure 1: Mean solution time in simulated and real group sessions

The solution time for the incentivized real groups is considerable faster than the constructed groups (drawing pairs form the individual incentivized session); 2.97 minutes versus 5.22 minutes. This difference is extremely improbable without synergies within teams; none of the 1000 samples have a mean successful solution time that is lower than than the mean from the true groups. For non-incentivized groups the mean solution time in the real group session is lower than the mean in the simulated sessions. But the differences is not significant as 201 out 1000 mean constructed solution time lies below the observed solution time in the groups session.

5 Conclusion

This paper brings home two messages. We replicate the experiment (Glucksberg (1962)) that seems to inspire the notion that monetary incentives impair individual creativity. In our data providing monetary rewards leave performance unaltered. This is also somewhat surprising, at least for those who think money incentives always induce individuals work harder and smarter. But it is a much less radical finding than the one reported in the Glucksberg experiment and which has been widely promoted by Dan Pink.

Our experiments indicate that there is a real group effect in creative problem solving, at least in the kind of "out of the box thinking" that is required to solve the Candle Problem. This is an important finding as it tells us something about how such problems ought to be organized. Furthermore it is not obvious that there would be a positive interaction effect in such a setting, as group work may be dominated by one person and hence bring fewer ideas on the table. It is important to investigate if the same effect

also appears in similar insight problems.

6 Appendix

Instructions (originally in Norwegian)

- *Group sessions (the part in brackets is information given only to the incentivized groups)*

Welcome to the experiment. (Depending on how quickly you solve a problem, you can each earn 1000 kroner or 200 kroner. The pair that solves the problem fastest earns 1000 kroner each. The pairs finishing in second or third positions earn 200 kroner each.) You work together to solve a problem and only the articles under the cloth may be used. The time will be measured. The time limit to solve the problem is 15 minutes. Please, indicate to experiment leader when you have understood the instructions.

- *Individual sessions (the part in brackets is information given only to the incentivized groups)*

Welcome to the experiment. (Depending on how quickly you solve a problem, you have the opportunity to earn 1000 kroner or 200 kroner. The participant that solves the problem quickest earns 1000 kroner each. Those finishing in the second and third positions earn 200 kroner.) You shall solve a problem and only the articles under the cloth may be used. The time will be measured. The time limit to solve the problem is 15 minutes. Please, indicate to experiment leader when you have understood the instructions.

- *Task description*

Attach the candle on the wall defined by the A4 sheet paper and make sure that the candle burns well and that no wax drips on the table or on the floor. The candle cannot be in contact with the table.

References

Bowles, S. and S. Polania-Reyes (2012). Economic incentives and social preferences: Substitutes or complements? *Journal of Economic Literature 50*(2), 368–425.

Charness, G. and M. Sutter (2012). Groups make better self-interested decisions. *The Journal of Economic Perspectives 26*(3), 157–176.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivations. *Journal of Personality and Social Psychology 18*(1), 105–115.

Duncker, K. (1945). On problem solving. *Psychological Monographs 58*.

Fehr, E. and A. Falk (2002). Psychological foundations of incentives. *European Economic Review 46*(4), 687–724.

Frey, B. S. and R. Jegen (2001). Motivation crowding theory. *Journal of Economic Surveys 15*(5), 589–611.

German, T. P. and M. A. Defeyter (2000). Immunity to functional fixedness in young children. *Psychonomic Bulletin & Review 7*(4), 707–712.

Gino, F., D. Ariely, et al. (2012). The dark side of creativity: Original thinkers can be more dishonest. *Journal of Personality and Social Psychology 102*(3), 445–459.

Glucksberg, S. (1962). The influence of strength of drive on functional fixedness and perceptual recognition. *Journal of Experimental Psychology 63*, 36–41.

Gneezy, U. and A. Rustichini (2000). Fine is a price, a. *Journal of Legal Studies. 29*, 1.

Holmstrom, B. and P. Milgrom (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics and Organization. 7*, 24–52.

Kugler, T., E. E. Kausel, and M. G. Kocher (2012). Are groups more rational than individuals? a review of interactive decision making in groups. *Wiley Interdisciplinary Reviews: Cognitive Science 3*(4), 471–482.

Maddox, W. W. and A. D. Galinsky (2009). Cultural borders and mental barriers: The relationship between living abroad and creativity. *Journal of Personality and Social Psychology 96*(5), 1047–1061.

Oyer, P. (1998). Fiscal year ends and nonlinear incentive contracts: The effect on business seasonality. *The Quarterly Journal of Economics 113*(1), 149–185.

Pink, D. H. (2010). *Drive: The surprising truth about what motivates us*. Canongate.

Department of Economics
University of Bergen
Fosswinckels gate 14
N-5007 Bergen, Norway
Phone: +47 55 58 92 00
Telefax: +47 55 58 92 10
http://www.svf.uib.no/econ