BILAG 49

Bilag 2

Statement of Dr. Anne Niedermann,
Institut für Demoskopie Allensbach

I, the undersigned, Dr. Anne Niedermann, Director of Research for Legal Evidence at the Institut für Demoskopie Allensbach, Germany, hereby declare as follows:

a. I am regularly asked by courts and private entities in several European countries to render expert opinions about fact-finding issues on the basis of surveys in connection with trademark disputes.

b. Along with a brief filed on August 19th, 2015, GLAXO SMITH KLINE introduced another survey conducted in Norway by GFK HEALTHCARE DIVISION in collaboration with the fieldwork provider, RESPONS ANALYSE AS (= Exhibit 16, entitled "272.201.10008 – Inhaler: Colour Association Survey – Norway", accompanied by an English translation of a statement by Mr. Morten Engan, the responsible researcher at RESPONS (= Exhibit 17. According to GSK's brief of 19 August, p. 4, this third GSK survey focuses on the perception of the colour shade PANTONE 2587C, which is one of the two colours featured in the GSK SERETIDE inhaler. The fieldwork took place between January and March 2015 (according to the statement by Mr. Engan = Exhibit 17, p. 3). I would like to point out that even though the data for the survey according to Exhibit 16 was established back in January-March of 2015, the translation of which was only brought to my attention on August 25th, 2015.

c. Furthermore, two statements, one by Prof. Lars Erling Olsen (= Exhibit 19) and one by Dr. Almut Pflüger (= Exhibit 18) pertain to this survey.

d. I was asked by the company SANDOZ to submit a statement on the probative value of the aforementioned third GSK survey in the pertaining case from my professional viewpoint as a social scientist specialised in survey research for legal evidence, including surveys on trade mark issues at both the national and the European level. The Allensbach Institute will charge SANDOZ for my work on this matter in accordance with the Institute's regular fee schedule but I render this statement in an independent, scientific and neutral manner on the basis of the attachments cited above.

I examined the RESPONS survey in the light of general scientific criteria as well as any applicable specific guidelines and rules developed by case law that should be adhered to for a survey to qualify as neutral, unbiased evidence with probative value.

With the third survey, GSK seeks to support a finding *of "distinctiveness of the colour purple, namely pantone 2587C"* (cf. brief of August 19th, 2015, p. 5). At the same time GSK refrains from deriving any conclusions from the survey with regard to the issue of a "mark with reputation".

My overall conclusion with regard to the method employed is the same as with the two former GSK-surveys (Exhibit 113 and 116): Also the survey submitted in the form of Exhibit 16 does not prove acquired distinctiveness in Norway.

My main reasons for coming to this conclusion are as follows:

In reaction to my criticism in my first statement on the first two surveys from Norway (= Exhibit 113 and 114 in the writ of 28 October 2014), the third GSK-survey omits several grave mistakes in the sampling and recruiting process that were detrimental to the probative value of both the first and second survey of 2014 (= Exhibit 113 and 114). It is correct to select professional respondents without any restrictions as to the number of inhalers sold or the frequency of prescribing them. Based on the available information, I will not raise objections as to the recruiting process here. I will also not attack the sample size, although the sample size of n = 145 among pharmacists and 150 among doctors is at a low end.

Nevertheless, implementing some lessons learned from earlier surveys, omitting some mistakes in sampling and analysis and making minor changes to the wording and order of questions still does not render the new approach a useful survey overall. Even the third GSK survey far from delivering a proper survey on distinctiveness. The persisting problems are not so much rooted within the technical execution by the fieldwork company RESPONS (sampling, recruiting etc.) but concern the covering of the whole relevant public as well as, most importantly, the test object and questionnaire wording.

1. Relevant public not covered entirely

A central criticism is that the third GSK survey (Exhibit 16) - in contrast to the two surveys before - completely lacks results for patients, even though GSK included this group in the first surveys as being just as relevant as the other two groups (doctors and pharmacists).

The problem of not being able to find enough respondents in Norway among patients that did not already take part in the two GSK surveys of 2014, is not inherent in the survey method as such. If GSK had commissioned an unbiased sampling approach from the very start and had refrained from conducting two surveys with biased method back in 2014 (= Exhibit 113 and 116), they would not

now have to deal with a shortage of respondents among patients in the new survey from Exhibit 16.

There is a basic requirement that must be fulfilled before a survey on the legal concept of distinctiveness can be afforded any probative value: the survey must have been conducted among the entire relevant public (which in this particular case also includes patients). In INTA's international review of rules for surveys that are submitted for legal evidence, the rule that the survey has to include "*all actual / prospective purchasers*" is included in INTA's list of recommended best practices.[1] In OHIM's Manual of Trade Mark Practice on surveys in the context of CTMs, the rule reads as follows: "*The criteria for the selection of the interviewed public must be assessed carefully. The sample must be indicative of the entire relevant public and, accordingly, must be selected randomly.*" (OHIM Manual, Part B, at 2.12.8.4). The respective rule is also emphasised in Norwegian sources, cf. LASSEN/STENVIK, Kjennetegnsrett, 2011, 244; VIKEN, Markedsundersøkelser som bevis i varemerke- og markedsføringsrett, 2012, 70; cf. also Swedish sources: NORDELL, Varumärkesrättens skyddsobjekt om ordkännetecknets mening och referens, 2004,197.

The definition of what groups of people altogether constitute a relevant public is basically a legal definition and is not part of the general rule illustrated above. The definition must be found in each case individually and may differ substantially from cast to case. Obviously, in the two surveys of 2014 (= Exhibit 113 and 114), GSK held the patients as relevant as doctors and pharmacists. If one regards patients are being part of the relevant public in the specific case at hand, one has to conclude that this third GSK survey (= Exhibit 16) violates the basic rule for survey evidence, namely that the survey has to cover the entire public. Because the third GSK survey does not give any results for patients (users), despite GSK regarded them a relevant group in the case at hand, the survey can be discounted as one-sided. It shows only one side of the coin, the perspective of professionals.

2. Test object

The test object as well as the question wording of the third GSK survey of 2015 (= Exhibit 16) are in large parts nearly the same as in the first one (= Exhibit 113), for

[1] INTA, Report on best practices in conducting surveys in trademark matters, 2013, para. I/C (http://www.inta.org/Advocacy/Pages/Reports.aspx).

which we know that the questions were drafted by GSK themselves. Not surprisingly, therefore, the third GSK survey does repeat severe mistakes, already highlighted by me in Exhibit 67 to the Defendant's brief of 9 December 2014 in relation to the test object and question wording, which are both at the core of the probative value of any survey - with the result that this third survey also does not constitute valid proof of the distinctiveness of the GSK SERETIDE appearance.

Moreover, the third survey does not reflect the fact that the GSK SERETIDE inhaler features a combination of two shades of colour on a three-dimensional shape: As in the previous GSK-survey (= Exhibit 113), the test object, a colour card, displays only one single colour shade whilst the question wording throughout the survey also does not pertain either to the issue of a combination of two shades of purple on the GSK SERETIDE inhaler, or, respectively, to the combination of two shades of purple plus the shape.

As in Exhibit 113 and throughout Exhibit 16 before it, the question wording does not even focus on the specific shade of purple claimed, as the question wording throughout the entire survey leads respondents to think unspecifically about the colour purple in general. We cannot know for sure whether each respondent was actually thinking about the particular colour shade at issue when answering or whether he/she perhaps had some other shade of purple in mind.

3. Test approach and questionnaire wording

The general test approach and the questionnaire wording employed by the GSK/RESPONS survey of 2015 (Exhibit 16) remain, once more, unclear. Neither the title of Exhibit 16 ("Colour association survey"), nor the accompanying statement of Mr. Morten Engan (Exhibit 17), the researcher at RESPONS responsible for conducting the fieldwork, identifies properly the legal concept the study was designed to measure. What Mr. Engan describes on p. 1 as the assignment: "*... to determine ... whether the (purple colour 'pantone 2587C ... has a clear recognition value with regard to inhalers...*" does not grasp the legal concept of acquired distinctiveness. Mr. Engan does not mention any specific personal experience in designing surveys for legal evidence either, rather he seems to be qualified as a general market researcher. This fact taken together with the further fact that he did not state that he himself developed the test approach as well as the questionnaire wording independently from GSK, and the fact that the wording of the core questions is basically the same as before in Exhibit 113—where we already know

that it was drafted directly by GSK—it becomes clear that the approach of the new survey in Norway was also not developed independently from GSK. This is confirmed by Dr. Pflüger's statement on the third GSK/RESPONS (Exhibit 18) on p.1: GSK themselves and the law firm Stephenson Harwood LLP wrote the questions, instead of the responsible survey researchers at GFK or RESPONS.

Prof. Olsen also does not name the legal concept that was researched by Exhibit 16. His description of the research topic as being "*to determine the association and perceived connection between the colour purple and GSK/Seretide*" (Bilag 19, p. 1) is not matching the legal concept of source distinctiveness for which GSK claims the results of the survey. Prof. Olsen defines furthermore the test object as being the basic "*colour purple*". This is imprecise. The colour purple per se includes many different shades. The legally relevant test object would rather be "a specific purple colour shade". "*Perceived connection*" is also unclear. What specific connection is GSK's survey all about?

Obviously, 3 out of 8 questions posed in connection with the test object (Q2, Q4 and Q8) are meant to establish a so-called "3-step test", a standard test for distinctiveness.[2] As for the Nordic countries, Sweden has established the same approach of a 3-step test, cf. PBR 05-080 (Kexchoclad). In the present case, the test has been incorrectly executed.

> The questionnaire includes more test questions than just the three core questions required for a normally straightforward and compact 3-step-test. There is one biasing introductory question, Q1, and there are more questions mixed in between the core questions (specifically Q3 and Q5). These additional questions are not neutral but diminish the validity of the intended test, since the core questions Q2, Q4 and Q7 are adversely affected by the additional prior questions. The additional questions induce socalled framing effects. It is important that the respondents of a survey on distinctiveness keep strictly concentrated only on the sign to be tested and come to the abstract decision whether the sign is from only one single source or not (that would be Q4 here) <u>before</u> they think about the market in general or about names of companies. Instead, in the present survey, their focus was redirected and they were led to think more broadly as early as Q1 and, again, by Q3 and Q5 just before the respective core questions Q2, Q4 and Q7.
>
> In light of this, Dr. Pflüger's conclusion regarding Q1 on p. 4 of her statement (Exhibit 18): "*Such extremely high correct attributions ... at the beginning of an interview are a first indication ... that there is a link between the colour shown in connection with inhalers and the company*" is on the one hand very true: It perfectly describes the

[2] E.g. on CTM-level R 1810/2008-4 – 3D mark shape of a suitcase / *RIMOWA III*, para. 35; R 355/2007-4 - Colour mark Orange and Grey / *Stihl*, para. 41 et seqq.; R 1/2005-4 – Red tool case / *Hilti*, para. 34 et seq.; T-164/03 [2005] - Monbébé, para. 80 et seqq.

market leader effect induced by Q1—which is not to be confused with source distinctiveness. On the other hand Dr. Pflüger's conclusion is definitely incorrect, as Q1 does not capture attributions (as Q4 should) but mere associations without any source exclusiveness. In Q1, respondents merely state what comes to mind when they see the card and hear the product category "*inhalers*" without any focus that could possibly be interpreted as measuring source distinctiveness.

In this very case, the share of awareness in Q2, which at first sight appears high, is easily explained by the frame of the prior question, Q1, which invites answers in line with the market leader effect: as soon as you think about the category inhalers (instead of sticking strictly to a focus on the sign to be tested), respondents are likely to remember and name market leaders, of which, of course, they are aware.

- It is important not to insert any questions in between the three core questions of a 3-step-test, especially not general questions, in order to avoid a market leader effect. In this case, a market leader of the product category inhaler in Norway is the one who claims distinctiveness (GSK). It is imperative, that market leaders take steps to prevent the probable market leader effect when conducting surveys on distinctiveness. That effect is when respondents name a particular trade mark or company (here: GSK or SERETIDE) just because the name of a market leader or its product comes to their mind in the form of a loose mental association when they think about the product category (inhalers) and not because the individual respondent actually connects the very sign in question (the colour shade) exclusively with the company or product of a market leader. In the third GSK survey, there were less questions in between the core questions than in the survey according to Exhibit 113, yet the questions Q1, Q3 and Q7 before the respective core questions still invited the market leader effect and, most probably, hyped the results in favour of GSK.

- The most severe mistake of the first GSK-survey (Exhibit 113) is repeated: again, the core question (in Exhibit 16 Q4) has far too imprecise wording.

  The need for precision in the question wording was pointed out in the Norwegian case (Supreme Court) Rt. 1979, 1117 (Cash & Carry). Imprecise questions lead to the conclusion that the survey has low probative value.

  In the present case, the report in English (= Exhibit 16) suggests that a certain share of interviewees selected the statement that reads: "*Inhalers of this colour originate from **one** (accentuation by AN) specific company*". In fact, the actual wording of the statement that was read out to the respondents was according to the questionnnaire (Exhibit 17, Question 4 on p. 6): "*Etter min oppfating, så kommer inhalatorer med denne fargen fra **et** spesifikt selskap*". Translated correctly into English, this would be "*In my opinion, inhalers with this colour come from **a** (accentuation by AN) specific company*". So, by choosing this statement respondents merely confirm the easy-to-agree-on obvious fact that each inhaler on the market naturally must have been manufactured by some company. The statement in the original Norwegian version does not convey source distinctiveness properly, as the necessary reiterated and

exclusive attribution to origin from only one single commercial source, which would be indicative of the distinctive character of a trade mark, is not captured. The wording of Q4 is unable to measure the core of the legal concept of source distinctiveness as exclusive attribution.

In light of this, it is misleading to the reader to describe in the report (Exhibit 16), e.g. in Summary B (table 6) and Summaries, as analysis of the group of respondents with "*awareness and indication of **Only** (sic!, accentuation by AN) one company*". "Only" is exactly the word, namely the necessary focus on source exclusivity, that is missing from the actual wording of the relevant statement in the core question Q4.

Dr. Pflüger is correct in mentioning that I follow the same basic approach that GSK is aiming for with Q4 (p. 4, fn. 4). However, the important difference in approach to the present GSK surveys (Exhibit 113 and 16) is that ALLENSBACH surveys use proper core questions that definitely capture source exclusivity.

Prof. Olsen seems to be of the opinion that the "*associations*" measured here instead of distinctiveness were helpful to clarify the legal concept to be proven. However, "*associations*" are irrelevant when it comes to source distinctiveness. Associations are a much more superficial type of mental connection than what is actually to be proven here: exclusive attribution in the sense of indication of origin to only one commercial source. Therefore, contrary to what Prof. Olsen highlighted on page 3 of Exhibit 19, the results on spontaneous associations as triggered by Q1 are of little evidential weight compared to results on source attribution (that would be the interrelated analysis of Q2, Q4 and Q7, see section 4 below). As far as Prof. Olsen is of the opinion that the associations measured would even prove a transfer of opinion to other products, this is not a relevant remark in the context of distinctiveness: Only if all doubts on the existence of original and/or acquired distinctiveness of a sign were removed, one could start to look at the degree of awareness of the test object, and only then associations come into play. But if, as is the case here, the discussion is still about the basics, the distinctiveness, it is not possible to turn already to the next step, the reputation or "well known status" of a trade mark that definitely has acquired distinctiveness.

- Ultimately, the third GSK survey submitted as Exhibit 16 includes no control group that could measure the strength of the aforementioned market leader effect in order to clean (subtract) it from the data. At least an unknown quantity of the responses that mention GSK or SERETIDE as the brand name or manufacturer in Q7 can partly be explained independently of the colour: be it because of a market leader effect (which is in the end a top-of-mind effect). The GCEU recognised this frequent deficit in surveys on acquired distinctiveness in which the market leader is involved in the decision "BIC".[3] In the present case, responses that name GSK or SERETIDE in Q7 may represent nothing more than a reflection of brand market share. The brand comes to mind the moment

[3] GCEU, T-262/04 [2005] – Bic, para. 84 et seqq.

respondents are told the category of product, namely inhalers in general, or they were just guessing. Therefore the GSK/GFK survey fails to establish true causality. It is not able to determine what share of responses that mention "GSK" and/or "SERETIDE" were actually caused by the shade of purple in question. For establishing true causality, a control group test is imperative and 'state of the art' in surveys for legal evidence when a market leader is involved, e.g. in the U.S. Control groups are also listed by INTA as recommended "best practice".[4]

In Prof. Olsen opinion a control group was not necessary. Clearly, providing the standard answering category "*don't know*" does not prevent guessing effects. And it is not a logically valid point to argue that the answers to Q4 (if the object "is from a specific company") would show that there is no adverse effect. This very question, Q4, might be inflated by the market leader effect and only a control group design can measure the exact share of guessing of a market leader effect. The control group is always set up to re-check exactly the core questions. Here that would be a check on Q4 as well as on Q2 and Q7. The opportunity to simply guess a market leader occurs inevitably in any study on a product that belongs to the market leaders from the very first survey question on, as already the mere mentioning of the product category or service category (here: "*inhalers*") triggers the undesirable effect of guessing a market leader within that product category, not necessarily the test object itself (here: the colour shade). The opportunity to guess is inevitably introduced by the necessity of naming the product category in the survey.

If survey results are challenged because of a market leader effect as in the present case, this is basically a challenge of the underlying causal proposition in cases such as this one where the owner of the sign is one of the market leaders: Is the test object actually the reason for the answers of the respondents or the product category itself and what exact share of the answers ground in general guessing? Acquired distinctiveness is only definitely proven if the extend of a probable market leader effect is determined by means of control group design and the percentage share of respondents that nevertheless name the respective brand or company even when seeing a completely different and non-confusable test object (here it would be another colour shade far from the colour violet) is subtracted from the survey results that relate to the colour shade in question.

- The problem that the GSK/RESPONS test approach does not properly reflect distinctiveness and that the answers are most likely mixed with loose associations to a market leader is confirmed by Q5 "*Why do you say that?*". If there were actual distinctiveness because of the colour shade (causality!), the answers to this question would accordingly clearly centre on the very colour shade tested as being the reason for the previous answer ("*from one specific company*") in

[4] Seidman Diamond, Reference guide on survey evidence, in: Federal Judicial Center: Reference manual for scientific evidence, Third ed., Washington (2011), 359-424, p. 397 et seqq.; INTA, Report on best practices in conducting surveys in trademark matters, 2013, para. I/C (http://www.inta.org/Advocacy/Pages/Reports.aspx).

Q4. Unfortunately, no results of Q5 are disclosed for the decisive group (respondents who (a) know the colour or are undecided if they know it, (b) attribute it at the same time to only one company, (c) name at the same time in Q7—not in Q1!—GSK or SERETIDE and (d) whose reasons in response to Q5 at the same time pertain to the specific colour shade). However, from the basic count provided for Q5 in table 7 one can at least obtain a rough picture: the percentage shares among all pharmacists and GPs that pertain to the very colour shade tested are definitely well below 50 percent: the result of a simple addition of the relevant answers[5]—ignoring for once the necessity of a net value count—is only 3 percent of all pharmacists and only 14 percent of all doctors.

4. Data analysis

There is a basic problem in connection with the data analysis by GFK/RESPONS in Exhibit 16. I recognise the effort of RESPONS to provide interrelated counts like I had postulated in my first statement on the GSK survey of 2014 (Exhibit 113). Nevertheless, the interrelated counts that provide summaries are all incorrect (tables 6, 10, 11, 12 and 13)[6]. They all integrate results pertaining to Q1 into the analysis and these are irrelevant in the 3-step test. Q1 captures associations of names that come to mind, not exclusive source attribution. Mere associations, however, are definitely irrelevant in an analysis of a 3-step-test. A proper analysis should have combined only Q2, Q4 and Q7. Especially the end result that is to be derived from table 6/Summary B is incorrectly established for this reason.

Dr. Pflüger glosses over this problem in the analysis several times in her statement: firstly by giving the diagrams on the right side of p. 3 that cite the values 84 percent and 65 percent from table 6/Summary B a title which suggests these were actually denoting the share of respondents that were "*Aware and indicating origin of one specific company Q4*". This is not the true picture because association results from the irrelevant Q1 were mixed into this analysis, as is revealed by the text above the matching table 6 of Exhibit 16 and an easy-to-miss hint in the left side of the diagram in Exhibit 18: Q1 is listed among the questions that were part of the analysis.

[5] "*GSK uses that colour/Seretide uses that colour*": 2 percent among GPs; "*I have only seen it on one type of inhaler/manufacturer*" 3 percent among pharmacists and 4 among GPs, "*generic manufacturers tend to use the original colour*" 2 percent among GPs; "*I think it is a patented colour/only Seretide GSK uses this colour/have not seen it on anything other than Seretide*" 3 percent among GPs; "*Would be confusing if other company used this colour*", 1 percent among GPs.

[6] The tables 11, 12 and 13 are irrelevant for the end result as they incorporate follow-up questions after Q4 which do not belong to the standard 3-step test (Q6, Q7).

Similarly imprecise is Dr. Pflüger's sentence on p. 5 suggesting that tables 8 and 9 infer that 56 percent of pharmacists name "...*GSK / GLAXO SMITH KLINE to be 'the one company'* " (on p. 5). Also into this analysis, results from Q1 were mixed in inappropriately which have nothing to do with attributing to only one company (cf. text above tables 8 and 9 of Exhibit 16).

The imprecision reoccurs when Dr. Pflüger cites on p. 6 and p. 8 of Exhibit 18 the figures from table 10/Summary C of Exhibit 16 that seems to specify the share of those who name a wrong manufacturer. In that case also, results from the irrelevant Q1 were mixed into the analysis.

To mix in association results from Q1 in any of the analyses might perhaps have been an attempt to heal the defective wording of Q4 as the wording does not capture exclusive source attribution. However, this is logically incorrect as Q1 does not presuppose exclusive attribution.

It is nevertheless informative to take a look at the results in table 6/Summary B:

To make an exception and include Q1 for once into the analysis, as no other values were disclosed, the 84 percent of all Pharmacists and 65 percent of all GPs is my starting point for a proper analysis. From Table 10/Summary C we know that 10 percent of the pharmacists and 6 percent of the GPs named the wrong company as the source. German case law always subtracts these shares from the share of respondents who know the sign in question and attribute it correctly to only one single source, based on EuGH, Slg. 2002, I-5475 Rn 65 = GRUR 2002, 804 – Philips (cf. BGH, GRUR 2007, 1066 Rn. 36 = WRP 2007, WRP Jahr 20071466 – Kinderzeit; BPatG, Beschl. v. 8. 3. 2013 – 33 W (pat) 33/12 = Vorlage zum EuGH zur Verkehrsdurchsetzung abstrakter Farbmarken – Sparkassen-Rot, para 3a) Allgemeine Grundsätze zur Ermittlung des Durchsetzungsgrades), even if social scientists have challenged the established practice (cf. NIEDERMANN GRUR 2006, 367, 371).[7] In the present case, in accordance with established practice, the actual end result based on the inflated analysis in table 6 would be for Pharmacists: 84 percent minus 10 percent = 74 percent, and the end result for GPs would be: 65 percent minus 6 percent = 59 percent. The Norwegian Court is free to use the deduction or to dismiss it.

If one takes into account the other problems of this survey explained above in sections 1 to 3, it is overall very unlikely that a proper survey (with a straightforward

[7] The German Federal Court of Justice, Dr. Pflüger and myself all agree that, in the context of measuring distinctiveness, it is <u>not</u> necessary that respondents were able actively to specify the correct name of the brand owner or the brand.

3-question-only test, with a valid core question that actually grasps exclusive attribution—even if one did not subtract the share of those respondents who attributed the sign to the wrong manufacturer as is prevailing OHIM practice—with a control group for determining and subtracting the market leader effect (as required by OHIM) would arrive at least among GPs (and, most likely also among the patients for whom we did not receive any new data by GSK) in shares clearly below 50 percent, thus failing the threshold national legislations throughout Europe apply to the 3-step-test and failing the requirement that sufficient distinctiveness must be present not only in some, but in every single relevant group. One cannot offset one group against the other or create an average value. This means that a higher degree of acceptance in one of the relevant groups cannot compensate for an inadequate degree in another relevant group.

I confirm that the above statement is truthful and my own.

Allensbach, 31 August 2015

Institut für Demoskopie Allensbach
Dr. Anne Niedermann